PROCESS CONTROLLABILITY ANALYSIS USING LINEAR AND NONLINEAR OPTIMISATION

Samara D. Chenery

October 1997

Submitted to the University of London
for the degree of
Doctor of Philosophy

Centre for Process Systems Engineering and
Department of Chemical Engineering,
Imperial College of Science, Technology and Medicine
London SW7 2BY

Abstract

Controllability analysis is concerned with determining the limitations on achievable dynamic performance. The information provided by this analysis can be used to screen initial designs and to make structural decisions while avoiding the expense of designing a control scheme. The objective of this work has been to develop such techniques for both linear and nonlinear process models, the results of which should be both unambiguous and easily interpreted giving a direct indicator of achievable performance.

This thesis proposes the use of optimisation techniques to determine the best achievable control performance for a system over a set of controllers for a specified set of disturbances.

Methods for solving such problems using both linear and nonlinear models have been developed and implemented using linear and nonlinear dynamic optimisation techniques. For the linear case an optimal control problem is formulated to assess the best achievable performance for the set of linear time invariant (LTI) controllers providing an upper bound on the controllability. This can be solved as a linear program (LP). While for the nonlinear case an optimal idealised control problem is formulated which provides a lower bound on the controllability. This is solved using nonlinear programming (NLP) and therefore provides a computationally expensive technique. Both of these bounds are developed to be as tight as possible. The nonlinear technique can be specialised to linear models and solved as an LP. This provides a quick and efficient result which can be used to provide a lower bound on the linear controllability.

These techniques provide a highly flexible framework for addressing typical process performance requirements through appropriate selections of the objective ,constraints and disturbance description. This framework is used to develop the OLDE and ONDE (Optimal Linear/Nonlinear Dynamic Economic) problems which provide controllability measures in terms of economic performance.

Acknowledgements

I would like to thank my supervisor Stephen Walsh for his support, encouragement and continuous guidance. I am also grateful to ICI for financial support of the project and the supply of industrial case studies.

I am grateful to all my friends and colleagues at the Centre for Process Systems Engineering for all their advice and suggestions and for making my PhD time enjoyable.

I also thank all my friends and family, particularly my parents, Karenza, Rachel and Anne, for their humour and their continued confidence in me.

Finally I'd like to express my thanks to Darryl Copsey for his patience and consideration.

Contents

List of Figures

Nomenclature

A_{ℓ_1}	matrix of 1's and 0's such that $(A_{\ell_1}\phi)_i = \sum_{j=1}^{n_w}\sum_{t=0}^{\infty}\phi_{ij}(t)$
A_q,h	the matrix and vector used to define the discrete elements of Φ in terms of the discrete elements of Q for the Q-approximation, i.e., $\phi + A_q q = h$
A_{zero}, b_{zero}	the matrix and vector defining the zero interpolation conditions, i.e., $A_{zero}\phi = b_{zero}$
$a_{pq}^{ij,\lambda_0,k}(l)$	elements of A_{zero}, the values of λ_0, i, j, k define which row the element is on, the values of p, q, l define which column
$\alpha_i(\lambda)$	ith row of $L_U^{-1}(\lambda)$ for $i = 1, 2, ...n_z$
$b^{ij,\lambda_0,k}$	elements of b_{zero} the values of λ_0, i, j, k define which row the element is on
$\beta_j(\lambda)$	jth column of $R_V^{-1}(\lambda)$ for $j = 1, 2, ...n_w$
c	the constraints
c_k	the kth constraint in c
$c_{nm}^{ij(k)}(p)$	elements of A_q the values of i, j, k define which row the element is on,the values of n, m, p define which column
c_z	number of constraints provided by the zero interpolation conditions
$\mathcal{D}$	open unit disc for λ
$\bar{\mathcal{D}}$	closed unit disc for λ
δ_i	maximum deviation of the ith constraint due to disturbances
δu	the deviation of u from u_0 used in Chapter 4
δw	the deviation of w from w_0 used in Chapter 4
δz	the deviation of z from z_{ref} used in Chapter 4

$E(o)$	the expected value of the objective o
$E(w)$	the expected value of the disturbance w, i.e., $\bar{w}$
ℓ_1	space of all ℓ_1 norm bounded operators, i.e., $Q \in \ell_1^{m \times n}$ if $\|R\|_1 < \infty$
ℓ_∞	the space of all infinity norm bounded signals, i.e., $\|z\|_\infty \leq \infty$
F	the function expressing the differential algebraic equations
$f(\mathrm{x})$	the objective function of a NLP
$G^{ss,cl}$	closed-loop steady-state gains
$G^{ss,ol}$	open-loop steady-state gains
$g(\mathrm{x})$	the inequality constraints of a NLP
g^p	the inequality constraints associated with the feasibilty subproblem for $w^p, \;\; p = 1, ..., n_{dist}$
$h(\mathrm{x})$	the equality constraints of a NLP
h^p	the equality constraints associated with the feasibilty subproblem for $w^p, \;\; p = 1, ..., n_{dist}$
h^{u_0}	the equality constraints enforcing a common operating point u_0
J	the objective function
J_o	optimal steady-state
J^o	the optimal dynamic objective
$\tilde{J}_1$	the objective function for selecting $\tilde{w}^o$
$\tilde{J}_2$	the objective function for ensuring infeasibilty for $\tilde{w}^o$ with intial control schedule $u_b^{\check{p},o}$
K	the controller
L	the length of the FIR of Q in the Q-approximation
$L_U M_U R_U$	Smith-McMillan decomposition of U
$L_V M_V R_V$	Smith-McMillan decomposition of V
$l^{\check{p}_j}$	the discrete time at which the jth component $w_j^{\check{p}}$ of disturbance $w^{\check{p}}$ steps, i.e., $l_j^{\check{p}} = t_j^{\check{p}}/T_{samp}$
l_k^{max}	the time of the maximum violation in constraint c_k
Λ_{UV}	set of zeros of U and V in $\bar{\mathcal{D}}$

λ	complex variable representing the unit delay $=z^{-1}$ for discrete z-operator, such that the λ-transform of the operator R is given by $R(\lambda) = \sum_{k=0}^{\infty} R(k)\lambda^k$.
λ_i	lagrange multiplier of constraints on z_i
N	order of the delay used for the Delay Augmentation algorithm
N_{fh}	finite horizon over which the linear specialisation of the nonlinear controllability technique is optimised
N, M	right-coprime factors of P_{22}
$\tilde{N}, \tilde{M}$	left-coprime factors of P_{22}
N_{ij}	finite horizon of ϕ_{ij} for linear controllability technique
$n_d ist$	the number of step disturbances in $\bar{W}$
ν_o	optimal solution
ν_{ij}	the deviation in z_i due to w_j
ν_i	the deviation in z_i due to all disturbances
ν^p	the maximum absolute deviation in z to each w^p
ν_i^p	the maximum absolute deviation in z_i due to w^p
$\nu_i^{p,+}, \nu_i^{p,-}$	the absolute maximum and minimum deviations in z_i due to w^p
$o(w,u)$	objective funtion for OLDE and ONDE techniques
$o(\bar{w})$	objective for expected disturbance
o^{ref}	reference point of o corresponding to w_{centre} and u_0
P	transfer function matrix from $(w \;\; u)^T$ to $(z \;\; y)^T$ for admissible systems
P_{22}	the actual plant
Φ	$H - UQV = H - R$, the objective function, a closed-loop map from w to z representing the performance objectives
Φ^+, Φ^-	sequences of maps with non-negative entries such that $\Phi = \Phi^+ + \Phi^-$
$\Vert\Phi\Vert_1$	ℓ_1-norm of Φ, i.e., $\Vert\Phi\Vert_1 = \max_{1\leq i\leq n_z} \sum_{j=1}^{n_w} \sum_{k=0}^{\infty} \vert\phi_{ij}(k)\vert$
Q	the stable parameter used to parametrise all stabilising LTI controllers
$sign_j^p$	the value to which the jth component w_j^p of disturbance w^p steps
$sign_j^{\check{p}}$	the value to which the jth component $w_j^{\check{p}}$ of disturbance $w^{\check{p}}$ steps

$sign_h^{\breve{p}j}$	the value to which the hth component $\tilde{w}_h^{\breve{p}j}$ of disturbance $\tilde{w}^{\breve{p}j}$ steps
σ_{U_i}	sequence of structural indices corresponding to U, ie, $\sigma_{U_i}(\lambda_0)$=multiplicity of the zero λ_0 as a root of the numerator of the ith diagonal term of M_U
σ_{V_j}	sequence of structural indices corresponding to V, ie, $\sigma_{V_j}(\lambda_0)$=multiplicity of the zero λ_0 as a root of the numerator of the jth diagonal term of M_V,
T_{samp}	the sampling period for the linear specialisation of the nonlinear controllability technique
$t_{d_{w_j,y_i}}$	time delay from disturbance component w_j to measurement y_i
$t_{d_{w_j,y}}$	minimum time delay from disturbance component w_j to any measurment y_i
t_j^p	the time at which the jth component w_j^p of disturbance w^p steps
$t_j^{\breve{p}}$	the time at which the jth component $w_j^{\breve{p}}$ of disturbance $w^{\breve{p}}$ steps
$t_h^{\breve{p}j}$	the time at which the hth component $\tilde{w}_h^{\breve{p}j}$ of disturbance $\tilde{w}^{\breve{p}j}$ steps
$t^p(k)$	the discrete times associated with $w^p(k)$, i.e., the disturbance has constant value $w^p(k)$ for the time period $t^p(k) \leq t < t^p(k+1)$
U	the set of acceptable control schedules , i.e., $u^l \leq u^p \leq u^h$
$\tilde{U}^{\breve{p}j}$	finite set of control schedules given by $\tilde{u}_b^p = u_b^{\breve{p},o}$ and $\tilde{u}_a^p = u_a^{p,o}$ for $p = 1, ..., n_{dist}$
u	control inputs, a vector of length n_u, in Chapter 3 this refers to the deviation from steady-state, in Chapter 4 this refers to the nominal value
$\bar{U}$	the finite set of selected control schedules $u^p,\ p = 1, .., n_{dist}$
u_0	the operating point
u_0^o	the optimal operating point
u_a	the control schedule $u(t)$ for $t \geq t_j^{\breve{p}}$
u_b	the control schedule $u(t)$ for $t < t_j^{\breve{p}}$
u^h	the upper magnitude bound on u
u^l	the lower magnitude bound on u

u_{lin}	steady-state of u about which linear model is linearised
$u^p \in U$	the control schedule selected for w^p
$u^p(k)$	the discrete elements of u^p
$u^{p,o}$	the optimal control schedule for disturbance w^p
$\tilde{u}^p \in \tilde{U}^{\check{p}j}$	the pth control schedule in the set $\tilde{U}^{\check{p}j}$
$\tilde{u}^{\check{p}j}$	the control schedule selected for $\tilde{w}^{\check{p}j}$
W_1	the filter to limit the rate of change of the disturbance
$\bar{W}$	the finite set of specified step disturbances
$\bar{W}^p$	the set of step disturbances for limiting the acausal behaviour with respect to disturbance $w^p \in \bar{W}$
$\tilde{W}$	the set of all possible $\tilde{w}_a^{\check{p}j}$, i.e., $\tilde{w}^{\check{p}j}(t), \;\; t \geq t_j^{\check{p}}$
w	disturbances, a vector of length n_w, in Chapter 3 this refers to the deviation from steady-state, in Chapter 4 this refers to the nominal value
w_0	the operating point of the disturbance w in Chapter 4
w_a	the disturbance $w(t)$ for $t \geq t_j^{\check{p}}$
w_b	the disturbance $w(t)$ for $t < t_j^{\check{p}}$
w_{centre}	reference point central to w^l and w^h
w^h	the upper magnitude bound on w
w^l	the lower magnitude bound on w
w_{lin}	steady-state of w about which linear model is linearised
$w^p \in \bar{W}$	the pth step disturbance in the set $\bar{W}$
$w^{\check{p}}$	a particular disturbance in $\bar{W}$
w_j^p	the jth component of disturbance w^p, $j = 1, ...n_w$
$w^p(k)$	the discrete elements of w^p
$\tilde{w}^{\check{p}j}$	the disturbance for limiting acausal behaviour with respect to the jth component $w_j^{\check{p}}$ of a particular disturbance $w^{\check{p}}$
$\tilde{w}_h^{\check{p}j}$	the hth component of $\tilde{w}^{\check{p}j}$
$\tilde{w}^o$	candidate disturbance for $\tilde{w}^{\check{p}j}$
$\bar{w}$	the expected disturbance

$\tilde{X}, \tilde{Y}$	stable part of the Bezout identity $\tilde{X}M - \tilde{Y}N = I$ for the right-coprime factorization
X, Y	stable part of the Bezout identity $\tilde{M}X - \tilde{N}Y = I$ for the left-coprime factorization
x	the optimisation parameters of a NLP
x_l,x_u	the lower and upper bounds on the optimisation parameters of a NLP
x^j	element of the extended set of right null chains of V, ie, $x^j = (x^j_1....x^j_{\sigma_j})$
y	measured outputs, a vector of length n_y
y^i	element of the extended set of left null chains of U, ie, $y^i = (y^i_1....y^i_{\sigma_i})$
y^{ref}	reference point of y corresponding to w_{centre} and u_0
z	regulated outputs, a vector of length n_z, in Chapter 3 this refers to the deviation from steady-state, in Chapter 4 this refers to the nominal value
$\|z\|_\infty$	infinity norm of z, i.e., $\|z\|_\infty = \sup_k \max_i |z_i(k)|$
z^h	the upper magnitude bound on z
z^l	the lower magnitude bound on z
z^{ref}	reference point of z corresponding to w_{centre} and u_0
z^{ss}_{ij}	the steady-state value of z_i due to w_j

Abbreviations

DA	Delay augmentation
DAE	Differential algebraic equations
FIR	Finite impulse response
ISE	Integral square error
LFT	Linear fractional transformation
LMI	Linear matrix inequality
LP	Linear program
LTI	Linear time invariant
MIMO	Multiple-input multiple-output
NLP	Nonlinear Program
NMP	Non-minimum phase

OLDE	Optimal linear dynamic economics
ONDE	Optimal nonlinear dynamic economics
PFR	Plug flow reactor
RHP	Right-half plane
SISO	Single-input single-output
s.t.	such that
SQP	Successive quadratic programming

Chapter 1

Introduction

Ziegler and Nichols (1943) first introduced the concept of controllability as a measure of "achievable control performance", taking into account the effect on this of the design of the process system itself, when they stated that :

"A poor controller is often able to perform acceptably on a process which is easily controlled. The finest controller made, when applied to a miserably designed process, may not deliver the desired performance. True, on badly designed processes, advanced controllers are able to eke out better results than older models, but on these processes there is a definite end point which can be approached by instrumentation and it falls short of perfection."

The aim of this work has been to develop means for quantifying this "end point", the best achievable control, unambiguously. Such quantitative measures of achievable performance are needed to screen out process designs and control structures for which performance requirements cannot be met.

Controllability analysis techniques should give unambiguous measures of the best achievable control performance prior to the design of a specific controller. If specific choices of the controller design, disturbance values and setpoint changes are made for a simulation then the resulting measure of performance will be biased by these choices. In this case it is difficult to discern whether the result is due to fundamental properties of the process that cannot be changed or simply due to the specific choices made for simulation.

In general a range of techniques is available with the useful ones forming a hierarchy of increasingly demanding (ie, better models, more computationally expensive), but increasingly discriminating tests. This hierarchy is represented pictorially in Figure 1.1.

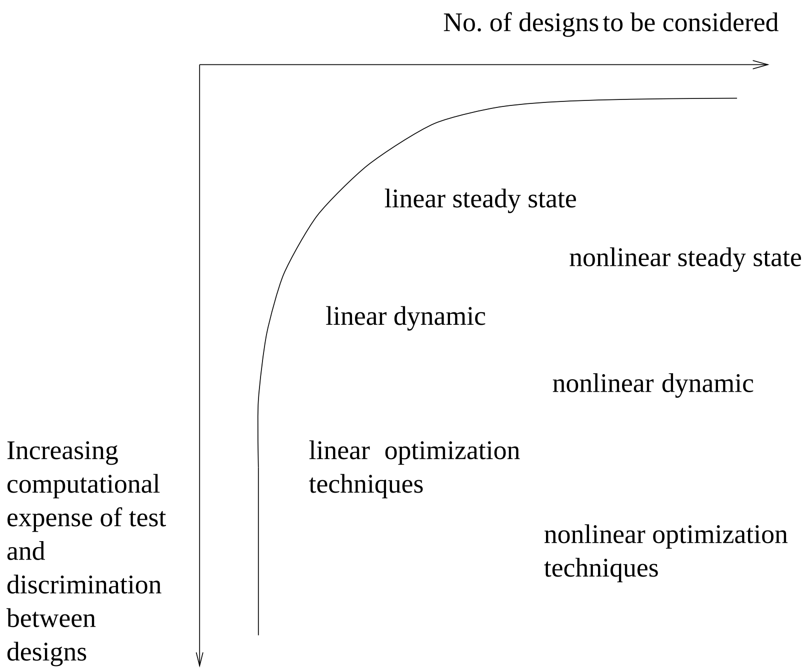

Figure 1.1: The control design "funnel"

It is the concept of controllability as "achievable control performance"that this thesis is interested in. However the term controllability has been given a range of definitions, such as, state controllability and functional controllability, therefore a brief discussion of these other concepts of controllability will be presented in section 1.1.

Techniques for assessing controllability (achievable performance) already exist for both linear and nonlinear models of the plant. The need for both linear and nonlinear analysis is discussed in section 1.2. Existing linear and nonlinear controllability techniques are discussed in Chapter 2. Section 1.3 presents a brief discussion of optimal control which we propose as a useful basis for new controllability analysis methods for both linear and nonlinear models. The methods developed for this thesis are presented in a general manner in section 1.4 and in detail in Chapters 3 and 4. This chapter concludes with a discussion of the structure of the thesis.

1.1 What is Controllability?

The definition of controllability is discussed in Skogestad (1994b). A brief summary is given here. In the 60's Kalman introduced the concept of state controllability which can tell us whether the state vector can be taken from any initial state to any final state within a finite time. However once all unstable modes have been shown to be controllable and observable this concept has limited practical significance. For example, although both initial and final states are defined, it is not possible to impose any conditions on the trajectory between these points or after the final time. Also it is not on the whole necessary to control and observe every state variable in the system. At this time the term "controllability" became widely associated with the narrower meaning of Kalman's "state controllability".

A further step in the development of a controllability concept which measured "achievable control performance" came when Rosenbrock (1970) showed that the invertibility of a system is a necessary and sufficient condition for functional controllability, where functional controllability is based on output reproducibility. Functional controllability tells us whether there exists, for a system, a set of input trajectories which can (with initial state zero) generate any output trajectory (which satisfies certain smoothness conditions). Therefore functional controllability of a system suggests that perfect control performance is possible. However functional controllability only provides a yes/no type answer and gives no measure of the achievable performance in the case when the system is not functionally controllable. Rosenbrock also introduced the notion of (RHP) right-half plane transmission zeros for multivariable systems which is now a fundamental part of controllability analysis.

Morari (1983) suggested that achievable control performance might be assessed independent of the controller. Such an assessment allows the consideration of questions about the ease with which a process might be controlled, the most suitable strategy for controlling it and changes that might be made to the process itself to improve control effectiveness. This method of controllability analysis allows the achievable performance of the plant to be assessed prior to the design of the controller. The choices of specific controller designs and specific values of disturbances and set point changes, made for simulations, bias the controllability assessment, such that it is unclear if results are fundamental properties of the plant or if they depend on the specific choices made. Morari suggested assessing how far a system is from achieving "perfect control" as a quantitative analysis of achievable performance. To avoid confusion with Kalman's "state controllability" he introduced the term dynamic resilience for describing controllability as a measure of the best achievable performance. However this term gives no suggestion of the concepts relation to control. In this thesis the term "controllability" will be taken as meaning "input-output (or output) controllability" for which Skogestad uses the following definition (Skogestad, 1994a):

Definition of (input-output) controllability: The ability to achieve acceptable control performance, that is, to keep the outputs (y) within specified bounds from their setpoints (r), in spite of unknown variations in the plant (e.g., disturbances (d) and model perturbations) using available inputs (u) and available measurements (e.g.,y_m or d_m).

From Wolff (1992) it can be taken that controllability and dynamic resilience imply the same thing, that is, a plants ability to achieve its specified control objectives.

A range of existing (input-output) controllability techniques are discussed in the next chapter. It is found that typically there is some ambiguity associated with these indictors and it is often unclear what the result indicates about the achievability of specific performance requirements. Therefore this thesis presents controllability analysis techniques, for both linear and nonlinear models, based on optimal control problems which provide unambiguous measures directly in terms of typical performance requirements.

1.2 Linear and Nonlinear Controllability

The measure of achievable performance provided by any controllability analysis is only as good as the model it uses. If the model is a poor representation of the actual process, then any conclusions made about the plants ability to meet certain performance requirements, which are estimated from it, will be erroneous. This tends to suggest that nonlinear models should be used, since they are more likely to capture the true behaviour of the plant than linear models. However linear analysis is greatly simplified by properties such as superposition. This means that the controllability analysis of the linear model of a plant is generally much simpler than that of its nonlinear model. If the plant is quite linear over the region of operation being investigated then linear controllability analysis can provide very useful results without too much effort (computational expense). Otherwise nonlinear controllability analysis will be necessary to ensure that any nonlinear behaviour of the plant is captured in the measure of achievable performance. Nonlinear controllability analysis is, as a rule, computationally expensive and therefore should be used after the linear analysis to validate the results.

1.3 Optimal Control

Although optimal control has been applied to the problem of controllability, it was originally set up, as a branch of modern control theory, for control design where the emphasis was on the controller selected rather than the performance achieved. The resulting system of any optimal control analysis is the system which gives the best achievable performance, as specified by an objective (cost) function and some constraints. Thus, with the objective and constraints chosen appropriately, optimal control gives a measure of the best achievable control performance which lends itself well to controllability analysis.

In Perkins and Walsh (1996) the attractions of such optimisation based techniques are given as:

"First, it has been shown that it is possible to devise absolute controllability tests by

this means, that is tests which if failed imply no real plant based on the tested partial design can meet the performance requirements. Second, it is possible to include the actual performance requirements (if known!) into the test itself"

This suggests that strong results, which directly relate to the performance requirements, might be produced by applying optimal control methods to the controllability problem, for linear as well as nonlinear models.

Many books exist which are purely dedicated to optimal control, for example (Whittle, 1996; Lewis, 1986; Bryson and Ho, 1975; Athans and Falb, 1966).

1.4 Thesis Work

The aim of this work has been to develop techniques which assess the best achievable control performance of a given plant. It is desirable to avoid biasing this measure by the specific choices of controller design or values of disturbances and set point changes made for simulations. The information provided by this analysis can be used to screen initial designs and to make structural decisions whilst avoiding the expense of designing a control scheme. The objective is to develop such techniques for both linear and nonlinear models, the results of which should be both unambiguous and easily interpreted giving a direct indicator of achievable performance. Therefore optimal control, which directly assesses the best achievable control performance of a system, can be used to provide an unambiguous result which directly indicates whether the system can achieve the performance requirements. The more general the set of controllers over which the optimisation takes place the closer the resulting measure will be to the true optimal control performance.

The controllability analysis developed in this thesis concentrates on the ability of a system to maintain performance requirements when subjected to a set of disturbances. If a controller cannot be found, in the set of possible controllers described for the problem, which makes the performance acceptable for all the disturbances in the disturbance set, then this suggests that the process is not sufficiently controllable for these performance requirements. In this case either the process must be redesigned or the requirements themselves be reconsidered. It follows that it is important to try to make the set of controllers, from which the optimal controller is selected, as broad and realistic as possible.

Techniques for the analysis of both linear and nonlinear controllability by solving optimal control problems have been developed and presented in this thesis (Chapters 3 and 4 respectively). The optimal control problem can be formulated with different objectives and constraints to give a range of performance measures. The problems can be set up to simply indicate feasibility, i.e., can the performance requirements be achieved with this plant, or to evaluate the optimal economic performance. The OLDE and ONDE (Optimal Linear/Nonlinear Dynamic Economic) problems provide a measure of the best achievable economic performance where the economics are represented, as in Narraway *et al* (1991), as the amount that the operating point must be backed off from the optimal operating point to ensure that none of the process disturbances cause violation of the process constraints.

The linear controllability analysis optimises the performance over the set of all stabilising linear time invariant (LTI) controllers, which is both a general and realistic set of controllers. This technique provides a pessimistic bound on the controllability. The disturbances can be selected from a combination of the set of all magnitude bounded persistent disturbances, step disturbances or steady state disturbances. The problem can be formulated and solved as a linear program (LP).

The optimal controller selected for the nonlinear controllability problem is idealised, providing an optimistic bound on the controllability. A technique for tightening this lower bound is presented. Since this bound is optimistic, if the process fails to be feasible for this test then it suggests it will be unable to achieve it with any real, implementable controller. The feasibility is tested over a set of user specified step disturbances. The linear controllability analysis can be used to provide good estimates of the worst step disturbance. The nonlinear controllability problem is formulated as a nonlinear programming (NLP) problem. This technique has been specialised to linear models, providing a computationally efficient lower bound on the linear controllability.

These two controllability techniques are designed to complement each other, with the linear controllability analysis being used initially to provide estimates of the achievable performance and the worst disturbance, and the nonlinear controllability analysis, which is much more computationally expensive, being used to validate the linear results.

If the process fails for the nonlinear controllability technique then this means that either, the nominal process is unable to meet the performance requirements with any controller, or the NLP has failed to find a global optimum. This gives a strong result, indicating that the process itself should be changed, before going to the expense of designing a controller that can never make the system achieve the specified performance. These techniques are useful as screening devices giving very direct and intuitive measures of achievable performance. Precisely where they should lie in the control design "funnel" depends on their computational cost. Since they provide both highly discriminatory results and are both computationally expensive, in that they require dynamic models and optimisation techniques, they should be placed towards the bottom of the funnel where most designs have been eliminated. The nonlinear controllability analysis is placed beneath the linear, since nonlinear optimisation techniques and the development of nonlinear models are much more computationally expensive than their linear counterparts.

1.5 Structure of Thesis

This thesis can be read at two levels. Chapters 1,2,5 and 6 provide a functional description of the motivation and results of the work. Chapters 3 and 4 complement this with detailed presentations of the linear and nonlinear controllability analysis techniques developed in this thesis.

Chapter 2 presents a critical review of the literature on controllability techniques. Controllability measures for linear models and nonlinear models are discussed and criticisms of their shortcomings presented. This chapter provides a motivation for developing unambiguous measures of controllability, for linear and nonlinear models, using optimal control. The aims of the work pursued in this thesis are discussed further.

Chapter 3 describes the proposed linear controllability methodology. The parametrisation of the controller, which allows the optimisation problem to be furnished with the set of all stabilising LTI controllers, and the feasibility constraints, which allow the problem to be formulated as a linear program (LP), are presented. The possible combinations of disturbance descriptions are explained and the range of performance requirements discussed. Finally some properties of the LP formulation and a brief overview of the implementation are presented.

The nonlinear controllability technique which has been developed is presented in Chapter 4. The nature of the idealised controller, its parametrisation and a technique for limiting the acausal nature of this controller are explained. The constraints imposing feasibility for each specified step disturbance are described and the application of a similar range of performance requirements as those for the linear controllability technique are discussed. Some details of the formulation and solution of the nonlinear program are given. A specialisation of this nonlinear technique to linear models, which is solved as an LP, is described. Finally a brief overview of the implementation is presented.

Chapter 5 contains a published linear example with which the linear software is validated, a couple of industrial case studies and a well published nonlinear problem on which the techniques developed in this thesis are demonstrated. Some practical shortcomings are discussed.

Finally the conclusion, Chapter 6, summarises how far the techniques developed in this thesis go towards answering the aims of the work, set out in Chapter 2. The contributions of this thesis are stated and some further work that might be undertaken in this area is suggested.

Chapter 2

Review of Controllability

This chapter describes a range of existing linear and nonlinear controllability techniques. The aim is to give a representation of the available techniques rather than an exhaustive review and the interest is primarily in controllability for full MIMO controllers rather than decentralised controllers. Further discussion of controllability analysis can be found in a book by Skogestad and Postlethwaite (1996).

Initially four fundamental limitations that prevent perfect control are described in section 2.1. The discussion of various linear controllability indicators is broken down according to these limitations on perfect control: RHP zeros, time delays, input constraints and model uncertainty. Both RHP zeros and time delays are non-minimum phase (NMP) characteristics. The nonlinear controllability measures are presented as analytical techniques and then optimisation based techniques. Linear optimisation based controllability techniques are included within the four categories of the linear controllability section with a discussion of general linear optimal control problems given in the next chapter. The short comings of the linear and nonlinear measures are discussed at the end of each respective section. Finally the motivation for the work in this thesis is presented.

2.1 Fundamental Limitations on Controllability

At present controllability analysis focuses on input-output controllability. The main objectives of a control system are to track the setpoint and to reject disturbances. The ideal controller accomplishes this by inverting the process.

In general for linear controllability analysis the process is modelled as a linear transfer function of the form

$$y(s) \;=\; G(s)u(s) + G_d(s)d(s). \tag{2.1}$$

where y are the measured outputs, u are the manipulated inputs, d are the disturbances, G are the plant model, G_d are the disturbance model, r will describe the reference inputs and C will describe the controller. The control error is given as $e = y - r$. Therefore the ideal controller would give a control error of zero by setting the manipulated input to the inverse of the process $u = G^{-1}r - G^{-1}G_d d$. In practise a feedback controller setting $u = C(s)(r - y)$ can accomplish something similar to this. The measured output and the manipulated input become,

$$y \;=\; Hr + SG_d d \tag{2.2}$$

$$u \;=\; G^{-1}Hr - G^{-1}HG_d d \tag{2.3}$$

where $S = (I + GC)^{-1}$ is the sensitivity function and $H = GC(I + GC)^{-1}$ is the

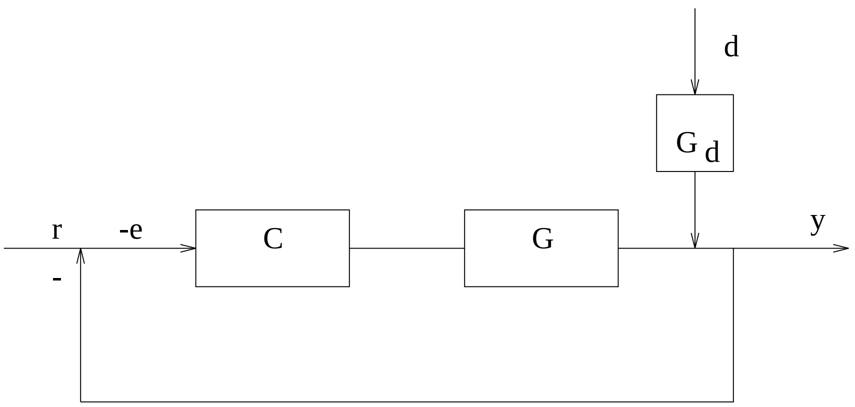

Figure 2.1: Block diagram of linear process model

complimentary sensitivity function. Perfect tracking and disturbance rejection require $S \approx 0$, $H \approx I$, which would give a controller output (2.3) corresponding to the inverse of the process. For this the magnitude of the loop gain, $L(jw) = GC(jw)$, must be much greater than one. Therefore control is generally only effective for frequencies less than a bandwidth frequency, w_B, which is defined as the frequency up to which the magnitude of the loop gain is greater than one, $\|L(jw)\| > 1$. Once $\|L(jw)\| \ll 1$ then $S \approx I$, $H \approx 0$. This suggests that ideal control requires high bandwidth, i.e., fast feedback.

There are four fundamental limitations that prevent fast control. These are:

- RHP zeros (a NMP characteristic),
- time delays (a NMP characteristic),
- constraints on the input variables,
- model uncertainty,

All these characteristics impose limitations on achievable bandwidth which conflict with the requirements for good control. This can be interpreted mathematically by the fact that each of these characteristics prevent the inversion of the process, since a right-half plane (RHP) zero is closely related to inverse response and implies an unstable inverse, time delays give rise to acausal elements in the inverse, input constraints and model uncertainty prevent the accurate inversion of the process. Therefore these prevent the implementation of the ideal controller. In practice measurement noise can also prevent the implementation of perfect control by inducing large variations in manipulated variables.

2.2 Linear Controllability

Controllability analysis for linear plants has been extensively explored and many controllability indicators have been developed. On the whole these only capture the limitation on the control performance for one of the fundamental limitations, mentioned previously, at a time. A range of such linear indicators for assessing the effect on controllability of NMP characteristics (RHP zeros and time delays), input constraints and uncertainty are presented in the following.

2.2.1 Non-Minimum Phase (NMP) Characteristics

The non-minimum phase (NMP) characteristics of a linear model are defined as the right-half plane (RHP) zeros and time delays of the plant G. There are a range of controllability indicators available that assess the controllability, independent of the controller, for RHP zeros and time delays separately. A technique that attempts to assess the best achievable performance for a linear SISO plant with both these NMP characteristics is the ideal integral square error (ISE) optimal control problem, described in Skogestad and Postlethwaite (1996). This involves the minimisation of

$$\mathrm{ISE} = \int_0^\infty |y(t) - r(t)|^2 dt \tag{2.4}$$

by the selection of the ideal controller $u(t)$. This controller is ideal in that there are no constraints on u included in the problem and the resulting optimal choice may not be implementable. The ideal response $y = Tr$ for a unit step in $r(t)$ has been shown (Morari and Zafiriou, 1989) to be

$$T(s) = \prod_j \frac{-s + z_j}{s + \bar{z}_j} e^{-\theta s} \tag{2.5}$$

where z_j is a RHP zero,θ is a time delay and $\bar{z}_j$ is the complex conjugate of z_j. Skogestad and Postlethwaite list the optimal ISE for a stable plant with a:

1. time delay θ as ISE=θ.
2. RHP zero z as ISE=$2/z$.
3. complex RHP zero $z = x \pm jy$ as ISE=$4x/(x^2 + y^2)$.

Techniques for assessing the limitation on controllability due to RHP zeros and time delays, separately, are presented in the following.

Right-Half Plane (RHP) Zeros

It has been well established that RHP zeros limit the achievable closed loop performance of both SISO and MIMO systems independent of the control system design. For example the presence of a RHP zero in a SISO system may give rise to inverse response behaviour and high-gain instability (Skogestad and Postlethwaite, 1996). According to Skogestad (1994b), which reviews fundamental results for controllability analysis of scalar systems, the upper bound on the bandwidth for a SISO system is given approximately by,

$$w_B \; < \; z/2 \text{ ,where z is a real RHP zero.}$$

The loop gain $|L| = |GC|$ drops below 1 at frequency $z/2$ and since $|L| > 1$ is needed for $S \approx 0$ and $H \approx 1$, which gives "perfect control", then plants with RHP transmission zeros within the desired bandwidth should be avoided.

Qiu and Davison (1993) extend the SISO ideal ISE optimal control problem to linear MIMO plants with RHP zeros at z_i. They show that the ideal ISE value for a step disturbance or reference relates directly to $2\sum_i \frac{1}{z_i}$, where the contribution for a conjugate pair of complex RHP zeros is the same as shown for the SISO case.

In Morari *et al* (1987) the concept of "zero-directions" is used to characterise the achievable transfer matrices. The paper uses the same block diagram as in Figure 2.1 to represent the feedback control system. It is assumed that the transfer matrix G(s) is square, has RHP zeros (z_i) of degree only one and has no RHP poles located at the RHP zeros (z_i). However it is suggested that these restrictions can be relaxed, although this would require a more involved notation.

Since z_i is a zero of G(s) then if G(s) has rank n, $G(z_i)$ has rank $n-1$, i.e., the RHP zero z_i causes G(s) to loose rank. λ_i is the zero direction associated with this zero as long as,

$$\lambda_i^T G(z_i) = 0, \qquad \lambda_i \neq 0. \tag{2.6}$$

The following transfer functions are useful for showing the effect of the zero direction.Let $G_d = I$ for,

$$T_{ur} \;=\; C(I+GC)^{-1} \tag{2.7}$$

$$T_{ud} \;=\; -T_{ur} \tag{2.8}$$

$$T_{yr} \;=\; GC(I+GC)^{-1} \;=\; GT_{ur} \tag{2.9}$$

$$T_{yd} \;=\; (I+GC)^{-1} \;=\; I - T_{yr} \;=\; I - GT_{ur}. \tag{2.10}$$

(2.6) taken together with (2.10) imply that

$$\lambda_i^T T_{yd}(z_i) = \lambda_i^T (I - G(z_i)T_{ur}(z_i)) = \lambda_i^T, \tag{2.11}$$

i.e., the magnitude of any disturbance d entering along the zero direction λ_i and passing through to the output y is unaffected by feedback. Thus RHP zeros (for MIMO) affect both the achievable T_{yr} and T_{yd}, where these are the transfer functions between the disturbance and the output and the reference signal and the output. Therefore plant RHP zeros will limit the achievable disturbance response despite the controller. The RHP zero z_i is "pinned" to the outputs corresponding to non-zero entries in λ_i and cannot affect those with zero entries. In fact if the jth entry of λ_i is the largest entry then z_i is said to be predominantly aligned with the jth output and to attempt to push it to another output would cause severe interactions. Hence any RHP zeros in the plant would ideally occur at high frequencies and have zero directions that are aligned with outputs which do not, for example, need good disturbance rejection.

Time Delays

A time delay has essentially the same effect as a RHP zero, in that it limits the achievable speed of the response by providing an upper bound on the bandwidth which for SISO systems is given in Skogestad (1994b) as,

$$w_B \;<\; 1/\theta \text{ ,where } \theta \text{ is a time delay.}$$

The analysis follows that of the RHP zero case with the low frequency asymptote of the loop gain $|L| = |GC|$ dropping below 1 at the frequency $1/\theta$.

In Holt and Morari (1985) simple procedures for assessing the best achievable response for MIMO systems independent of the controller are discussed in terms of the effect on controllability. Holt and Morari transform the structure in Figure 2.1 to give the equivalent basic internal model control (IMC) structure in Figure 2.2 to represent the linear model of the system. No generality is lost and it is more useful for the analysis of controllability.

In IMC "perfect control" can be achieved directly by setting $G_c = G^{-1}$. In the case when we want to concentrate on the analysis of delay times we assume a perfect model with unlimited controller power and a stable inverse (no RHP transmission zeros) therefore an inability to implement the inverse would indicate that G contains predictive elements due to time delays, ie , G^{-1} is not causal. In this case G is factored, $G = G_+G_-$, where G_-^{-1} is stable and causal and G_+ is noninvertible. The controller is then chosen as $G_c = G_-^{-1}$, $G_+(0) = I$ and G_+ is the closed loop transfer function of the system.

Since G_+ cannot be uniquely defined the goal of the paper is to find the optimal G_+, i.e., that which will minimise the integral square error (ISE) and the integral absolute error (IAE) as defined in the paper (Holt and Morari, 1985).

Holt and Morari (1985) provide an upper bound on the controllability (dynamic resilience) without dynamic decoupling by giving a lower bound for the settling time of each output i.

$$\tau_i = \min_j p_{ij} \tag{2.12}$$

p_{ij} = minimum time delay in numerator of element ij of G.

The results are given in a matrix factor G_* which might require the addition of off-diagonal elements or an increase in the delay of the diagonal elements to make it a valid choice for G_+.

$$G_+ \ = \ G_* = \begin{bmatrix} e^{-\tau_1 s} & * & . & . & * \\ * & e^{-\tau_2 s} & & & \\ . & & . & & \\ . & & & . & \\ * & & & & e^{-\tau_n s} \end{bmatrix} \tag{2.13}$$

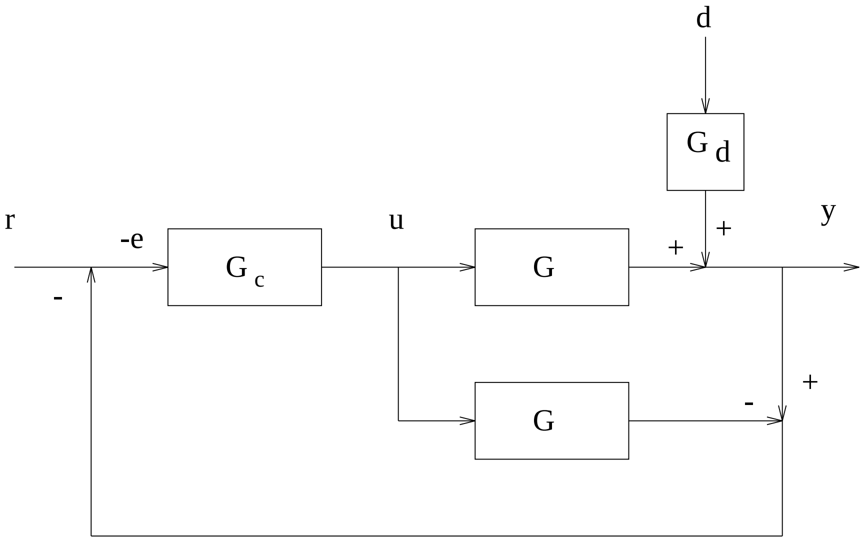

Figure 2.2: Block diagram of IMC structure

They then give a lower bound on the controllability (dynamic resilience) by giving the minimum time delay for each output when the system is dynamically decoupled.

$$G_+^d \ = \ diag(r_{11}, ..., r_{jj}, ..., r_{nn}) \tag{2.14}$$

with

$$r_{jj} \ = \ \exp(-s(\max_i(\max(0, (\hat{q}_{ij} - \hat{p}_{ij}))))),$$

$$\hat{p}_{ij} = \text{minimum delay in numerator of element ij of } G^{-1},$$

$$\hat{q}_{ij} = \text{minimum delay in denominator of element ij of } G^{-1}.$$

The matrix G_+^d is realizable and any controller (with or without decoupling) which has any outputs with greater delay times is not optimal in the sense mentioned earlier.

By consideration of the diagonals of both G_+ and G_* we have a measure of the controllability (dynamic resilience), since the smaller the minimum response time (ie the time delays in the diagonals) the better the controllability. However it is not obvious how to judge what is a reasonable diagonal for good controllability.

Perkins and Wong (1985) suggest that these measures for assessing controllability are awkward to use and that a scalar measure of the impact of delays on controllability would be preferable. The paper discusses the assessment of achievable plant performance, specifically the effect of time delays, using the concept of functional controllability which was mentioned in the introduction. Their results are presented for the discrete case, although they state that similar results hold for the continuous case. Functional controllability and the notion of "perfect control" are related by the invertibility of the plant. In fact the invertibility of G(z) is a necessary and sufficient condition for functional controllability.

Theorem(Rosenbrock, 1970): Given a transfer function matrix G(z) with McMillan degree p and a sequence of outputs 0,0,..,0,y_p,y_{p+1},..., then there exists a sequence of inputs u_0,u_1,... which generates the output sequence given x_0=0 if and only if det[G(z)]$\neq$0.

The period of time p (the McMillan degree) during which the output does not change is Rosenbrock's attempt at defining the time delay for the MIMO system. However this measure of the minimum necessary time delay is given by the degree of the monic least common denominator of all minors of all orders of the matrix and is generally a gross overestimate.

They present a theorem which calculates a scalar measure of the minimum time delay, allowing the independent specification of all outputs, and from this they state that it is possible to devise a generic algorithm to compute the minimum delay which can be used to establish some simple bounds on its value. Psarris and Floudas (1990) point out that this minimum time delay is also given by the largest element, r_{jj}, from the diagonal of Holt and Morari's minimum time delay matrix G_{+}^{d}. The problem of what is a reasonable value of the minimum necessary delay, i.e., "what is too large?", still remains.

Russell and Perkins (1987) broach the problem of giving controllability analysis a physical interpretation in the time domain, specifically relating their approach to time delays. They suggest the use of structural controllability for state-space systems as an approach which does not rely on an input-output model and which is based in the time domain. A system represented by $(\underline{A}_0, \underline{b}_0)$ is structurally (state) controllable if and only if there exists a completely state controllable system that can be represented by $(\underline{A}, \underline{b})$, which are structurally equivalent to $(\underline{A}_0, \underline{b}_0)$. A structural matrix has only two types of entries, fixed zeros (these can never take non zero value) and arbitrary entries (these may take any value including zero). This representation is proposed to avoid the difficulties of deriving a numerical model of the plant and the biasing of controllability analysis by the choice of parameters describing the system.

They then go on to outline an approach to controllability based on the analysis of the structural matrices of differential-algebraic equation (DAE) systems extended to time delays. They rank control schemes for a system by their generic minimum necessary delay:

Definition: The minimum necessary delay τ_{ming} of a system with time delay matrix $\underline{D}$ is the lowest minimum necessary delay τ_{min} which can be achieved by any system with the same delay matrix $\underline{D}$.

The delay matrix $\underline{D}$ is incorporated into the DAE system so that there is a non-negative delay associated with each occurrence of a variable in an equation. The structural matrix for this system is augmented with columns and rows for potential control structures. This matrix must be structurally nonsingular for τ_{ming} to be finite. Cause and effect paths are identified for different control schemes and the minimum necessary delay evaluated. The control schemes can be ranked by their respective generic minimum necessary delays. This mean that this analysis of controllability is control structure dependent, ie, the evaluation of τ_{ming} is used to compare control schemes rather than evaluate limitations on achievable performance inherent to the process itself.

2.2.2 Input constraints (actuator limits)

If disturbances or reference inputs cause the manipulated input to violate input constraints then the plant is no longer invertible, i.e., the manipulated input $u = G^{-1}r - G^{-1}G_d d$ cannot take its required value, therefore the ideal controller cannot implement the inversion of the process and "perfect" control is not possible. Therefore bounds that prevent such constraint violations and measures of the effect of disturbances or reference signals on the input are of interest as controllability indicators.

Many controllability measures consider the magnitude of acceptable deviations in which case it is convenient to scale the variables. On the whole the variables are scaled so that they should lie within the interval -1 and 1, i.e., their desired or expected magnitudes should be normalised to be less than 1 for all frequencies. A discussion of scaling can be found for the SISO case in Skogestad (1994b) and for the MIMO case in Wolff *et al* (1992).

For a SISO system we have the input constraint $|u(jw)| \leq 1$ $(\forall w)$, so to fulfill this requirement and track reference signals ($|r(w)| = 1$ for $w < w_r$), Skogestad (1994b) gives a lower bound on the transfer function G ,

$$|G(jw)| > 1 \ \ \forall w < w_r. \tag{2.15}$$

For MIMO systems the input required for tracking a sinusoidally varying reference signal $r(jw)$ is given by (Wolff *et al.*, 1992) as,

$$\frac{1}{\overline{\sigma}(G)} \leq \frac{\|u\|_2}{\|r\|_2} \leq \frac{1}{\underline{\sigma}(G)}. \tag{2.16}$$

where $\overline{\sigma}(G)$ and $\underline{\sigma}(G)$ are the maximum and minimum singular values of G respectively. Since r may have any direction, a small $\underline{\sigma}(G)$ implies that large input magnitudes might be needed. Therefore a small $\underline{\sigma}(G)$ is undesirable since it might require the violation of input constraints to maintain tracking. However judging a good value for $\underline{\sigma}(G)$ relies heavily on correct scaling, i.e., regarding a value of $\underline{\sigma}(G)$ of less than 1 as too small could be misleading if u and r are imprecisely scaled.

When it comes to disturbance rejection, Skogestad (1994b) gives a lower bound on the transfer function G for a SISO system which must perfectly reject a disturbance (for $w < w_d$) whilst maintaining $|u(jw)| \leq 1$ ($\forall w$),

$$|G(jw)| \; > \; |G_d(jw)| \;\; \forall w < w_d. \tag{2.17}$$

where $|G_d| > 1 \; \forall w < w_d$. Note that if the system is unstable with a real RHP pole at p then we need

$$|G(jw)| \; > \; |G_d(jw)| \;\; \forall w < p. \tag{2.18}$$

where p may be larger than w_d. For MIMO systems the input magnitude needed for perfect rejection of the worst disturbance ($\|d\|_\infty \leq 1$) is given by (Wolff *et al.*, 1992) as,

$$\|u\|_\infty \; = \; \|G^{-1}G_d\|_\infty \tag{2.19}$$

which is the largest row sum of the absolute values of the elements of $G^{-1}G_d$. Therefore a frequency dependent plot of the elements of $G^{-1}G_d$ will provide insight into the possibility of violating input constraints and which disturbances are most likely to cause such problems. The above gives constraints involving G and G_d which must be satisfied for perfect control with input constraints, similar constraints for acceptable control can be described (Skogestad and Postlethwaite, 1996) but are not discussed here.

Another indicator of the effect of input constraints on controllability, based on assessing the magnitude of the control input required to reject disturbances, is the disturbance condition number which was introduced by Skogestad and Morari (1987b).

$$\gamma_d(G) \; = \; \frac{\|G^{-1}g_d\|_2}{\|g_d\|_2}\,\overline{\sigma}(G) \tag{2.20}$$

where g_d is used instead of G_d to indicate that only one disturbance is considered at a time. The disturbance condition gives the ratio of the magnitude of u required to perfectly reject a disturbance in the direction of g_d to that required to reject a disturbance of the same magnitude, but in the "best" direction. The "best" direction is the direction that requires the least control action. If $\gamma_d(G)$ is large it suggests a large increase in the input magnitude which might again require the violation of input constraints to maintain disturbance rejection. $\gamma_d(G)$ is scale dependent and it is not obvious what is meant by too large.

Other such controllability techniques, i.e., the relative disturbance gain (RDG), the closed-loop disturbance gain (CLDG), etc, for assessing the achievable disturbance rejection performance with input constraints are discussed in Skogestad and Wolff (1996).

2.2.3 Model Uncertainty

Model uncertainty prevents the accurate inversion of the plant. It requires the controller be detuned and performance be sacrificed.

Measures of the effect of model uncertainty on controllability (dynamic resilience) are discussed at length in a paper by Skogestad and Morari (1987a). The structured singular value $\mu(M)$ was considered the best measure:

$$\text{robust stability iff } \mu(M) \leq 1 \; \forall \, w,$$

where M is the interconnection matrix describing the nominal transfer functions from the output of the perturbations to their inputs for the $M\Delta$-structure used for robust stability analysis. However at the time it was concluded that since it required a control system to have already been designed it was unsuitable for "screening purposes at the design stage". However μ-synthesis is not dependent on the selection of a controller design since it searches for the optimal controller that minimises this μ-condition. This μ-optimal control problem cannot be directly solved, however a popular technique for tackling this problem is DK-iteration which is described in Skogestad and Postlethwaite (1996).

For independent uncertainty in the elements of the plants transfer function matrix the interconnection matrix M can be rewritten as $LC(I+\tilde{G}C)^{-1}E = L\tilde{G}^{-1}\tilde{H}E$, where

$\tilde{G}$ and $\tilde{H}$ are used to denote the nominal plant and complementary sensitivity functions respectively (note that y = Hr), and L and E are weighting functions. In this case upper bounds for $\tilde{H}$ can be devised for uncorrelated element uncertainty (Skogestad and Morari, 1987a)

$$G - \tilde{G} = E\Delta L, \quad \Delta = diag(\Delta_{ij}), \quad \overline{\sigma}(\Delta_{ij}) < 1.$$

with the nominal response decoupled with identical responses such that $\tilde{H} = \tilde{h}I$, i.e.,

$$\text{robust stability iff } \|\tilde{h}\| \; < \; \frac{1}{\mu(L\tilde{G}^{-1}E)} \;\; \forall\, w. \tag{2.21}$$

Since $H \approx I$ implies perfect control this provides a limit on the achievable control for uncorrelated element uncertainty based only on plant information and the weights L and E.

In Wolff *et al* (1992) the condition number $\gamma(G)$ and the relative gain array (RGA) are presented as useful measures with respect to element uncertainty, but only if the relative errors of the transfer matrix elements are independent. Skogestad and Morari (1987a) go further to show that the usefulness of the RGA and the condition number in assessing the impact of uncertainties is limited not only to element uncertainties that are independent, but also to those with similar relative magnitude bounds.

The RGA is scale independent and useful as a measure of the effect of uncertainty on controllability, because its entries, i.e.,the relative gains λ_{ij}, provide an indication of how sensitive the plant is to uncertainty in an element.

$$\lambda_{ij} \; = \; g_{ij}(s)[G^{-1}(s)]_{ji} \tag{2.22}$$

If the element $g_{ij}(jw)$ were to vary by $-1/\lambda_{ij}(jw)$, i.e., $\tilde{g}_{ij}(jw) = g_{ij}(jw)(1-1/\lambda_{ij}(jw))$, then $G(jw)$ would become singular, therefore large RGA-elements imply that the $G(jw)$ may become singular for small relative perturbations in certain elements of the transfer function matrix.

The condition number is described by Perkins and Wong (1985) as,

$$k(G) = \|G\|.\|G^{-1}\|, \tag{2.23}$$

where $\|G\|$ is any matrix norm. $k(G)$ gives an upper bound on the ratio between the relative error in u and the relative error in G.

$$\frac{\|\delta u\|}{\|u\|} \leq k(G).\frac{\|\delta G\|}{\|G\|} \tag{2.24}$$

To remove the effect of scaling, from the comparison of k for different designs, they suggest optimally scaling k

$$k_{min}(G) = \min_{D_O,D_I} k(D_O G D_I), \tag{2.25}$$

,where D_O and D_I are diagonal matrices with real, positive entries. This scaling problem has been solved by Bauer (1963) for both the induced 1-norm ("max column sum") and the induced ∞-norm ("max column sum").

For the induced 2-norm the condition number becomes the ratio between the largest singular value $\overline{\sigma}(G)$ and the smallest singular value $\underline{\sigma}(G)$,

$$\gamma(G) = \frac{\overline{\sigma}(G)}{\underline{\sigma}(G)}. \tag{2.26}$$

For this case the minimised condition number $\gamma(G)^*$ is given by

$$\gamma^*(G) \;=\; \min_{D_O,D_I} \gamma(D_O G D_I) \tag{2.27}$$

and the minimised absolute condition number is described in Skogestad and Morari (1987a) as:

$$\gamma_a^*(G) \;=\; \min_{D_O,D_I} \frac{\overline{\sigma}(|D_O G D_I|)}{\underline{\sigma}(D_O G D_I)} \tag{2.28}$$

A large condition number implies an ill-conditioned plant.

The minimised absolute condition number can be used to give an upper bound on the diagonal elements of $\tilde{H}$ (=diag($\tilde{h}_i$)), assuming the nominal response is decoupled, (Skogestad and Morari, 1987a).

$$g_{ij} = \tilde{g}_{ij}(1 + r_{ij}\Delta_{ij}), \;\; \|\Delta_{ij}\| < 1, \;\; r_{max} = \max_{ij} r_{ij}.$$

Robust stability if

$$\|\tilde{h}_i\| \;<\; \frac{1}{r_{max}\gamma_a^*(\tilde{G})} \;\; \forall\, w \,\forall\, i. \tag{2.29}$$

It can be shown that (2.29) is conservative when the relative uncertainties on the elements r_{ij} are different. Skogestad and Morari state that this bound is tightest for a 2x2 plant with equal r_{ij} ($\tilde{h} = \tilde{h}_i \ \forall i$). They then go on to say that the RGA gives an upper bound on the absolute minimised condition number and, in fact, that for real matrices and high condition numbers the value of $\|RGA\|_1$ approaches γ_a^*. This suggests that, for guaranteed robust stability, ill-conditioned plants with large RGA should be avoided since even when the relative model uncertainty r_{max} is small the performance in terms of $|\tilde{h}|$ would be very limited.

Skogestad and Morari (1987a) conclude that the minimised and absolute minimised condition numbers are only reliable indicators of the sensitivity to element uncertainty if the uncertainty is independent and norm-bounded, with similar relative error bounds.

Skogestad and Havre (1996) draw conclusions on what the input minimised condition number, condition number and RGA imply about robust performance in the presence of input uncertainty. Input uncertainty is described by

$$G = \tilde{G}(I + E_I) \tag{2.30}$$

where if $E_I = diag\{\epsilon_1, \epsilon_2, ...\}$ we have diagonal input uncertainty, while if all the elements of E_I are nonzero then we have full block input uncertainty. Skogestad and Havre (1996) make the point that " diagonal input uncertainty is *always* present in real systems". In this case the minimised condition number is given by

$$\gamma_I^*(G) \ = \ \min_{D_I} \gamma(GD_I) \tag{2.31}$$

They summarise that if the condition number $\gamma(G)$ is small it suggests robust performance to both diagonal and full input uncertainty. A small minimised condition number $\gamma_I^*(G)$ only indicates robust performance to diagonal input uncertainty. Finally a plant which gives large RGA elements will be fundamentally difficult to control.

2.2.4 Criticism of Linear Controllability Analysis Techniques

All the papers discussed so far rely heavily on linear models. Although this probably will not matter for systems operating within one operating region, if the system moves from one region to another then the linearised model may no longer be suitable. Also many of the techniques, such as the minimum time delay techniques, the input magnitude $\|u\|_\infty$ required for perfect disturbance rejection, the condition number $k(G)$ and the RGA, assume square plants G(s), however it is not realistic to believe that all control systems have the same number of inputs as outputs.

The analysis presented by these papers takes place in the frequency domain, but all the performance requirements and the set of disturbances are typically defined in the time domain and, on the whole, are impossible to represent precisely in the frequency domain. Therefore the analysis of controllability in the frequency domain presents some problems.

There is also the problem of relating these theories to the physical system, so that conclusions might be drawn about suitable design changes to overcome problems revealed by the analysis. This has been broached in Russell and Perkins (1987) in which it is blamed on the use of the input-output model and the frequency domain representation.

The applicability of measures of the effect of uncertainty in the model on controllability are limited to specific uncertainty descriptions and therefore their usefulness is restricted. For example, the usefulness of both the RGA and condition number are limited to uncorrelated element uncertainties with similar relative magnitude bounds.

Another problem associated with the constraints is the definition of the disturbances experienced by the system. On the whole the system is scaled, so that d is bounded with a norm less than or equal to one. For a SISO system Skogestad (1994) describes the disturbances as sinusoidal with magnitude $|d| \leq 1$. For MIMO systems the d is simply described in Wolff *et al* (Wolff *et al.*, 1992) as "several" disturbances all with $\|d\|_\infty \leq 1$. The general problem is that the effect of d is only measured as its worst norm effect and although frequency weighting can be applied to $\|d\|_\infty \leq 1$ the disturbance description cannot be restricted arbitrarily.

Further to these problems is the ambiguity in interpreting the results of all these controllability analysis techniques. What is a reasonable time delay and time delay diagonal G_+^d? What is too large a condition number? What is too large an RGA-element? etc. Using these tools of controllability analysis would require both the use of heuristics to simplify the task and experience. Morari (1992), notes that "we often do not quite understand yet when and how these indicators should be used", but even so "they are applied widely and indiscriminately and lead to erroneous conclusions about controllability". The ambiguity involved in the interpretation of these measures is a major problem for controllability analysis. One reason for the ambiguity is that each indicator typically only considers one of the control performance limitations, i.e., RHP zeros, time delays, input constraints or uncertainty, in isolation. Another is that they do not relate directly to performance specifications since, as mentioned earlier, these are generally time domain, whilst most of the linear indicators are frequency domain.

Therefore there is a need for an easily interpreted linear controllability analysis technique that handles non-square control systems, deals directly with time domain performance requirements, allows disturbances to be defined in the time domain and quantifies the combined effect of as many fundamental limitations as possible.

2.3 Nonlinear Controllability

Linear controllability is a well established field with a broad range of indicators for assessing how close to perfect control a plant lies. However these techniques require models that have been linearised about a steady state. Hence if the plant is highly nonlinear or the analysis is required for a new steady state then these linear results can no longer be relied on.

Therefore controllability analysis techniques that can be applied to the nonlinear model of the plant directly are required. The existing nonlinear controllability analysis techniques can be divided into those which are analytical and those which are optimisation based.

2.3.1 Analytical Techniques

A selection of nonlinear controllability analysis techniques are based on the concept of functional controllability, which relates to whether a set of input trajectories u exist that will generate any output trajectory y. The terms (right) invertibility, functional controllability and functional reproducibility are all used to express the same concept. Perfect control is achievable only if a system is functionally controllable. A series of papers have undertaken to establish conditions for the functional controllability of a nonlinear system. Sufficient conditions for the functional controllability of nonlinear systems were introduced in Hirschorn (1979) the theory of which was improved on by Singh (1982). Tsinias and Kalouptsidis (1983) then presented necessary and sufficient conditions for the functional controllability of single input systems. Finally necessary and sufficient conditions for the functional controllability of general multivariable systems described by

$$\begin{aligned} \dot{x} &= f(x,u) \;\; x(t_0) = x_0, \;\; x \in R^n, \; u \in R^m \\ y &= h(x,u) \;\; y \in R^r \end{aligned} \tag{2.32}$$

were derived in Li and Feng (1987). The main criticism of any technique based on functional controllability is that it only indicates a yes/no type answer. In the case when the system is not functionally controllable no measure is given of just how far from perfect control the achievable performance is.

A fundamental limitation on perfect control for linear systems is the presence of right half plane transmission zeros which give rise to an unstable inverse. However for nonlinear systems we do not have zeros and poles as such. The analogous problem for a nonlinear system is unstable zero dynamics. Zero dynamics and inversion for nonlinear systems are discussed in Daoutidis and Kravaris (1991) where a nonlinear system is described as minimum phase if its (unforced) zero dynamics is asymptotically stable and non-minimum phase if its (unforced) zero dynamics is unstable. The (unforced) zero dynamics of a nonlinear system is given by the dynamics of its reduced inverse, which is a minimal-order realisation of the systems inverse. Therefore an investigation of the zero dynamics of a nonlinear system will reveal whether a nonlinear system has a stable inverse or not. However once we have assessed that a system has an unstable inverse this will not identify the actual best achievable performance for this nonlinear system.

The relative order of a system has been introduced (Daoutidis and Kravaris, 1992; Soroush, 1994) as an analysis tool for the structural evaluation of alternative control configurations. Daoutidis and Kravaris (1992) interpret the relative order of a nonlinear system, described by

$$\begin{aligned} \dot{x} &= f(x) + \sum_{j=1}^{m} g_j(x)u_j + \sum_{k=1}^{p} w_k(x)d_k, \quad x \in R^n \\ y_i &= h_i(x) \quad i = 1, ..., m \end{aligned} \tag{2.33}$$

as the "structural analog of dead time". It quantifies such notions as "initial sluggishness", "direct effect" and "physical closeness". The relative order r_{ij} of output y_i with respect to input u_j is given by the smallest integer for which

$$L_{g_j} L_f^{r_{ij}-1} h_i(x) \neq 0 \tag{2.34}$$

or ∞ if no such integer exists. The relative order ρ_{ik} of output y_i with respect to disturbance d_k is given by the smallest integer for which

$$L_{w_k} L_f^{\rho_{ik}-1} h_i(x) \neq 0 \tag{2.35}$$

or ∞ if no such integer exists. $L_f h_i(x)$ is the Lie derivative given by $\sum_{l=1}^{n}(\frac{\partial h(x)}{\partial x_l}) f_l(x)$, higher order Lie derivatives are given as $L_f^k h_i(x) = L_f L_f^{k-1} h_i(x)$ and mixed Lie derivatives are given as $L_{g_j} L_f^k h_i(x)$. Daoutidis and Kravaris show that the relative order r_{ij} captures the aspects mentioned above for the response of output y_i to input u_j and similarly ρ_{ik} captures these aspects for the response of output y_i to disturbance input d_k. The smaller this relative order the more direct the effect, the less sluggish the initial response and the smaller the deadtime. Therefore the smaller the relative order

r_{ij} the better. In Daoutidis and Kravaris (1992)the relative order matrix is introduced

$$M_r = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & & \vdots \\ r_{m1} & \cdots & r_{mm} \end{bmatrix} \tag{2.36}$$

to capture the structural coupling between manipulated inputs and measured outputs. When possible the matrix is rearranged so that the minimum relative order for each row lies on the diagonal, thus suggesting the pairing of input u_i with output y_i, and allows the interaction of these input/output pairings with the other inputs and outputs to be investigated through the off-diagonal elements. The technique is extended to nonlinear discrete systems in (Soroush, 1994). This method allows the coupling and interaction for specific control structures to be assessed, however it does not indicate whether specific performance requirements might be achievable or not.

In a similar manner the relative gain array (RGA) is used for linear systems to assess the interaction of input/output pairings. It is also used as a measure of the effect of uncertainty on controllability. Mijares *et al* (1985) show how the steady state RGA can be calculated for general nonlinear systems. It is shown that the formulas used for computing the nonlinear RGA are of the same form as those used for the linear case. The block relative gain (BRG) is extended to nonlinear systems in Manousiouthakis and Nikolaou (1989), giving the steady state NBRG and dynamic DNBRG. The linear BRG provides information on the limitations on achievable decentralised closed-loop performance, as well as the restrictions on the robustness characteristics of this closed loop. The DNBRG is developed to provide a measure of the interaction between decentralised feedback loops for dynamic nonlinear systems. It is controller independent, assessing the closed-loop performance using only plant information, i.e.,

$$\begin{aligned} DNBRG_l &= N_{11}(N^{-1})_{11} \\ DNBRG_r &= (N^{-1})_{11}N_{11} \end{aligned} \tag{2.37}$$

where the plant is given as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = N(u_1, u_2)$$

so that $y_1 = N_1(u_1, u_2 = 0) = N_{11}(u_1)$, where u_1 and y_1 are the inputs and outputs which we are considering interconnecting using feedback control, and N^{-1} is the inverse of the plant such that $u_1 = (N^{-1})_1(y_1, y_2 = 0) = (N^{-1})_{11}(y_1)$. If the resulting DNBRG$= I$ then it suggests that the feedback controller connecting these inputs and outputs can be designed independently of the controller that will be implemented on the rest of the inputs and outputs, u_2 and y_2, since there is no interaction. Therefore the distance of the DNBRG from the identity operator I is suggested as a measure of the effect of the second loop on the first loop. The DNBRG is scale dependent which complicates the issue of how to interpret the values of the results. The steady state NBRG is shown to be a lower bound for the condition number of a nonlinear system. It is not clear for either the nonlinear steady state RGA or the NBRG/DNBRG what the results they provide indicate about the best achievable control performance and whether it can satisfy certain performance specifications.

2.3.2 Optimisation Based Techniques

Approaches for nonlinear optimisation problems are reviewed in (Perkins and Walsh, 1996; Walsh and Perkins, 1996). Some approaches which are specifically interested in assessing nonlinear controllability are summarised in the following.

Much work has been done on developing optimisation problems to assess performance under plant uncertainty. Grossman *et al* (1983) set up a feasibility problem in which the optimiser attempts to meet all the process constraints with an idealised controller over the set of uncertain parameters as follows

$$\begin{aligned} \chi(d) &= \max_{\theta \in T} \min_{z \in Z} \max_{j \in J} f_j(d, z, \theta) \\ \theta \in T &= \{\theta | \theta^L \leq \theta \leq \theta^u\} \\ z \in Z &= \{z | z^L \leq z \leq z^u\} \end{aligned} \tag{2.38}$$

where the process constraints are described as $f(d, z, \theta) \leq 0$, d are the design variables, z are the operating variables and θ are the uncertain parameters. This work was extended by Swaney and Grossman (1985) to assess the flexibility of a process by maximising a scalar, called the flexibility index F.

$$\begin{aligned} F = \quad & \max \delta \\ s.t. \quad & \max_{\theta \in T(\delta)} \min_{z \in Z} \max_{j \in J} f_j(d, z, \theta) \leq 0 \\ & \delta \geq 0 \\ & T(\delta) = \{\theta | \theta^N - \delta \Delta\theta^- \leq \theta \leq \theta^N + \delta \Delta\theta^+\} \end{aligned} \tag{2.39}$$

This index gives a measure of the greatest deviation in the uncertain parameters for which the design can maintain feasibility. However both these techniques only assess the feasibility and flexibility of a process operating at steady state. Thus Dimitriadis and Pistikopoulas (1995) extended these ideas to present a systematic framework for assessing both the flexibility and feasibility of dynamic systems.

Mohideen *et al* (1996) go on to present an optimal control problem based on the dynamic feasibility problem presented in (Dimitriadis and Pistikopoulos, 1995), but with an economic objective, as well as the feasibility condition $f(d, z, \theta) \leq 0$, which accounts for the capital costs. The objective is to select the design variables and control scheme to minimise the cost (of process units, controllers and operating) while remaining feasible over the finite time horizon under both parametric uncertainty and process disturbances. In this paper the control scheme is defined as multiloop PI-controllers which means that the result is a pessimistic bound on the optimal control performance in that the set of controllers, from which the optimal controller can be selected, has been restricted.

A steady state measure for quantifying the ability of a heat exchanger network to cope with inlet and target temperature changes was presented in (Saboo *et al.*, 1985). This measure, the resilience index (RI), is a similar concept to the flexibility index which gives a measure of the maximum range of uncertainty that a plant can handle while remaining feasible. The RI is defined as the maximum total disturbance load that a network can allow without becoming infeasible.

$$\mathrm{RI} = \max_{z,l} \sum_i |l_i| \tag{2.40}$$

subject to a set of system constraints on the system variables z and the disturbance load vector l. This problem can be generalised to any nonlinear system with disturbances l and system variables z. Unlike the flexibility index the RI cannot describe an asymmetric range, however finding the RI is much less computationally expensive.

Walsh and Perkins (1992) provide an optimistic bound on the disturbance rejection performance of a plant when limited by both time delays and uncertainty. This technique assesses the performance for an idealised controller under worst case conditions. The controller is ideal in so far as it is allowed to act perfectly once the minimum time delay has passed. The minimum time delay for each constrained output is made up of the time delay between the disturbance and the measured variable $t_{d_{d,y_j}}$ and the time delay between the manipulated variable and the measured variable $t_{d_{u_i,y_j}}$ and is given by

$$t_{d_j} = \min_{j \in J_d} \{ t_{d_{d,y_j}} + \min_{i \in I_j^s \cap I_j^c} \{ t_{d_{u_i,y_j}} \} \} \tag{2.41}$$

where $y_j | j \in J_d$ is the subset of measured variables that can usefully detect the disturbance, $u_i | i \in I_j^s$ is the subset of manipulated variables which have a strong enough effect on y_j and $u_i | i \in I_j^c$ is the subset of manipulated variables which are permitted to be connected to y_j in the restricted control structure. The optimiser is used to select the worst possible combination of process disturbances and uncertain parameters. The optimisation problem solved is based on an open loop dynamic version of the feasibility problem (2.38)

$$\begin{aligned} \exists d \quad s.t. \quad & \max_{\theta \in T} \max_{k \in K} \max_{t \in [0, t_{d_k}]} f_k(d, \theta, t) \leq 0 \\ & T = \{ \theta | \theta^L \leq \theta \leq \theta^u \} \end{aligned} \tag{2.42}$$

where t_{d_k} is the minimum time delay for the kth constraint. Similarly a dynamic open loop version of the flexibility problem in (2.38) can be solved to quantify the degree of infeasibility.

The effect of fundamental plant characteristics on its ability to cope efficiently when moving between different operating point, its switchability, was assessed by White (1994) as an optimal control problem. In this problem the optimiser selects the best switching trajectory for the plant, as well as time-invariant design parameters which allow for modifications of the plant design to improve switchability. Path constraints are incorporated to enforce magnitude bounds on the constrained variables and end-point constraints are used to ensure the plant reaches its new steady state by the end of the problems finite time horizon. Vu, Bahri and Romagnoli (1997) solve an optimal control problem which incorporates operability into the switchability problem.

2.3.3 Criticism of Nonlinear Controllability Analysis Techniques

There are notably fewer analytical techniques for nonlinear controllability than for linear controllability. Those that have been discussed give either a yes/no type result or a result which is difficult to interpret.

The yes/no type result is given in response to the questions of functional controllability and stable zero dynamics. When the answer is no we are left with no idea of just what the achievable performance might be and whether it might satisfy the performance requirements. Even when the answer is yes the measures have not incorporated all the possible limitations on control performance, particularly input constraints and uncertainty, and therefore it is still not clear that perfect control would be achievable in the presence of these also.

For the results which are difficult to interpret, i.e. the relative order matrix, the nonlinear RGA and nonlinear BRG, a general guide-line, whether an element should be big or small, for the desirable result is known. However it is not obvious what is considered too big or too small a value of an element of any of these arrays. On the whole, as for the linear controllability analysis indicators, this requires experience to decide. Also, whether we decide a result indicates good controllability or not, it does not give a specific estimate of the achievable performance which makes it impossible to assess whether the performance requirements for the plant might be achievable or not.

None of these analytical techniques incorporate all the fundamental limitations on controllability and the majority only incorporate one at a time.

The optimisation based techniques give results which are easily interpreted and relate directly to the performance constraints specified for the plant, since on the whole they are set up to optimise the achievable performance in terms of these constraints.

The majority of fundamental limitations, i.e., uncertainty, input constraints, unstable zero dynamics, dead bands, can be captured by these optimisation techniques, particularly if a feedback controller, such as a PI controller, is implemented. However if an idealised controller is used then generally these techniques do not incorporate any of the limitations on achievable performance associated with the measured variables available or the causal nature of a feedback controller.

As mentioned above these techniques tend to involve the optimisation of the performance by the selection of an optimal control. The nature of this controller affects the type of bound that this measure gives on the best achievable control performance. For these optimal control techniques to give a realistic measure of the best achievable control performance then the optimisation should be furnished with the complete set of implementable controllers. When the optimisation set of controllers is a restricted, but realistic set, i.e., the set of PI controllers (Mohideen *et al.*, 1996), then the result gives a pessimistic bound. When the optimisation set of controllers is an idealised set, i.e., as in Walsh and Perkins (1992), then the result gives an optimistic bound. Therefore, although these results are easy to interpret and relate directly to performance constraints, there is a degree of ambiguity in assessing just how realistic these bounds are.

A final criticism of these approaches to optimisation based nonlinear controllability must be the computational expense involved in the solution of such nonlinear dynamic optimisation problems.

In spite of the computational expense of these optimisation based techniques and the nature of the bounds they produce, they offer a very attractive solution to the controllability analysis technique. They provide unambiguous results which relate directly to performance requirements, allow constraints and disturbances to be defined in the time domain and can be formulated to incorporate more than one of the fundamental limitations on controllability within the framework of one analysis technique.

2.4 Motivation for thesis

From the reviews of linear and nonlinear controllability analysis it can be seen that controllability analysis techniques can be split into:

- analytical techniques
- optimisation techniques.

Although the linear controllability indicators have been categorised by the fundamental limitation that they measure, they can also be split into those based on optimal control, i.e., the ideal ISE optimal control problem (time delays and RHP zeros for SISO or just RHP zeros for MIMO), μ-synthesis (uncertainty), and those not.

The discussion of these linear controllability indicators led to the conclusion that there is a need for a linear controllability analysis technique which provides an unambiguous measure of the best achievable performance, handles time domain constraints and disturbances and incorporates the majority of fundamental limitations on controllability.

A look at the existing nonlinear controllability techniques revealed nonlinear optimisation techniques which give controllability results that relate directly to the performance specifications, allow the analysis to take place in the time domain and can be formulated to handle more than one of the fundamental limitations on controllability at a time. However these techniques are computationally expensive and provide optimistic and pessimistic bounds on the best achievable performance depending on the description of the optimisation control set. To make these bounds tight ideally this set should be as broad and realistic as possible.

Therefore the analytical indicators have been dismissed for the purpose of this thesis. However the optimisation based techniques seem promising and it is this branch of controllability analysis that is pursued in this thesis.

2.4.1 The Aim

The aim of the work in this thesis has been to find controllability analysis methods for linear and nonlinear models, based on optimal control problems, which satisfy as many of the following as possible:

- capture typical time domain performance requirements and disturbance descriptions directly;
- find the best performance with as broad and realistic a set of controllers as possible;
- take account of as many of the fundamental limitations on controllability as possible simultaneously;
- are computationally tractable.

2.4.2 Optimal Realisable Control vs. Optimal Idealised Control

Any controllability analysis technique based on optimal control can take one of two forms:

- optimal realisable control
- optimal idealised control

The first form of optimal control means that the controller could actually be implemented between a set of measured variables y and manipulated variables u, e.g., a stabilisable PI or LTI controller. This includes feedforward, as well as feedback, controllers since these simply involve adding the disturbances to the measured variables. Examples of this type of optimal control for the linear case are $\mathcal{H}_\infty$ optimal control which is used in μ-synthesis, $\mathcal{H}_2$ optimal control which includes LQG control and ℓ_1 optimal control. Unless the optimisation is furnished with the complete set of realisable controllers, this type of technique will give a pessimistic (upper) bound on the optimal achievable control performance, since the set of controllers is restricted and may not include the optimal controller. This type of optimal control problem has been used in this thesis to provide a measure of linear controllability analysis by selecting the optimal linear time invariant (LTI) controller. This technique is based on ℓ_1 optimal control theory, which is a well researched and existing area and is presented in detail in a book by Dahleh and Diaz-Bobillo (1995). ℓ_1 optimal control is used as a basis for this controllability technique, since it goes a long way towards answering the aims set out in section 2.4.1. This motivation is discussed in more detail in 3.1. However to provide the corresponding measure for nonlinear models involves tackling a nonlinear optimisation over the set of LTI controllers. To attempt to parametrise this set will produce a high dimensional nonlinear optimisation problem. Also the problem will no longer have the convexity properties of the linear case and hence global optimality of solutions cannot be guaranteed. Therefore optimal realisable control was only used to develop a linear optimisation based technique which is presented in detail in Chapter 3.

The second form of optimal control, optimal idealised control, means that the controller is idealised in the sense that it is optimal for the problem and may not be implementable. Examples of this type of optimal control are the ideal ISE optimal control problem or the minimum time delay optimisation problem given by Walsh and Perkins (1992). This type of technique provides an optimistic (lower) bound on the optimal achievable control performance, since the optimal controller may not be in the set of realisable controllers. Such an optimal control method has been used to develop a nonlinear controllability analysis technique and a specialisation of this technique to linear models. For the nonlinear optimisation there is still no guarantee of a global optimum, but there is better experience with the solution of such optimal idealised control problems, than with the nonlinear version of the optimal LTI control problem. Both the nonlinear controllability analysis technique and its linear specialisation are described in Chapter 4. The majority of the chapter is dedicated to the nonlinear method, whilst section 4.8 is used to clarify some differences in the formulation of the linear specialisation which allow it to be solved efficiently as an LP.

Therefore three controllability analysis methods have been developed in this thesis, a linear controllability analysis method based on optimal LTI control, a nonlinear controllability analysis method based on an optimal idealised control problem and its linear equivalent. All these techniques use discretised controllers, with both the linear methods requiring discretised linear models. This means that the discretisation method and sampling period are important to the techniques. Also they have been developed to meet as many of the desired characteristics for a controllability analysis method, stated in 2.4.1, as possible. However one area which is not tackled here is the incorporation of the fundamental limitation on achievable control performance due to uncertainty in the model. This area is avoided since these problems are computationally expensive as it stands and the addition of uncertainty would add further computational difficulty, making it even harder to tackle realistic problems. Also in reality it is hard to give an accurate definition of the uncertainty set, therefore to incorporate uncertainty would introduce conservatism and ambiguity to the measures. However this is an area that is open to further research as is suggested in Chapter 6.

Chapter 3

Linear Controllability Using Optimisation

The following chapter presents a technique, which provides a measure of the best achievable control performance of a linear model for a linear time invariant (LTI) controller, by solving a linear optimal realisable control problem. This measure provides a pessimistic bound on the actual best achievable performance since it does not include all the possible realisable controllers. For example a PI with output limiting is not included in this description. However since the set of all LTI controllers is a very general set this measure is not too biased by the choice of a specific controller design. It takes account of the majority of the fundamental limitations on controller performance mentioned in Chapter 2 , excluding model uncertainty. It also captures a wide range of typical performance requirements, depending on its formulation, and allows the direct inclusion of time domain constraints. The results are easily interpreted and the technique lends itself to use as a feasibility test, i.e., can performance requirements be met for this process design.

This controllability technique uses linear programming techniques to address the linear optimal control problem:

$$\min_{K,u_0} J(K,u_0) \;\; s.t. \;\; c(K,u_0,w) \leq 0 \; \forall w \in W \tag{3.1}$$

where K is a linear time invariant (LTI) controller and the objective function J and the constraints c are expressed as linear functions so that the problem can be solved by linear programming. u_0 is the operating point and $w \in W$ are the disturbances. This formulation encompasses a wide range of optimal control problems ranging from minimising the maximum deviation in objective variables subject to disturbances of magnitude less than one (the ℓ_1 optimal control problem) to optimising the expected value of a linear economic objective (the Optimal Linear Dynamic Economics – OLDE – problem).

The general development of the field of linear optimal control is presented and the use of the theory of ℓ_1 optimal control as a basis for this controllability analysis technique is motivated.

Using the appropriate controller parametrisation means that the achievable closed loop transfer functions for a generalised plant are expressed as an affine function of a stable parameter Q. This allows the achievable discrete time closed loop impulse responses to be defined by an infinite set of linear constraints and approximated by a finite set of linear constraints.. The methods for doing this are well established (Boyd *et al.*, 1988; Boyd and Barratt, 1991; Dahleh and Diaz-Bobillo, 1995) and are discussed in section 3.3.

The disturbance set W can be expressed in a flexible manner and can include a mixture of magnitude bounded signals, step signals and steady-state signals. The allowable combinations of disturbances can be restricted if desired. This flexibility is important to allow specific process requirements to be captured with adequate precision. We show how the effect of this general set W can be captured in a linear program.

Possible objective functions are then discussed and it is shown that many objectives of interest can be captured in a linear program.

Having defined the LP problems of interest we then discuss the properties of alternative formulations.

3.1 Linear Optimal Control

Linear optimal control involves the solution of an optimal control problem where both the plant and the controller are linear. Over the past decade there have been considerable advances in linear optimal control strategies, relative to the linear quadratic gaussian (LQG) methodology which was dominant in the 60's and 70's. This method, discussed in more detail in Anderson and Moore (1989), involves the optimisation of LQG models where these models have linear dynamics, the measurement noise and disturbances (process noise) are gaussian stochastic processes and the objective (cost) function is quadratic. However, as stated in Skogestad and Postlethwaite (1996), due to its reliance on accurate models and its assumption of white noise disturbances it was found that "LQG designs were sometimes not robust enough to be used in practise".

These short comings of LQG control led to the development of $\mathcal{H}_\infty$ robust control. It was found that, in spite of the lack of robustness in the LQG being attributed to the form of the $\mathcal{H}_2$ norm in the objective, both $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control had many similarities. $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control involve the optimisation of the respective, $\mathcal{H}_2$ and $\mathcal{H}_\infty$, norms of a closed loop transfer function matrix, given by the lower linear fractional transformation described in section 3.2. The LQG problem is a special case of $\mathcal{H}_2$ optimal control. The $\mathcal{H}_\infty$ approach can be combined with uncertainty representations to provide robust performance problems, i.e., μ-synthesis.

Another development in this field was the discovery that linear matrix inequalities (LMIs) could be used to reduce a wide range of problems that arise in system and control theory to standard convex optimisation problems (Boyd and Barratt, 1991; Boyd *et al.*, 1994). Such problems include $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems. However solving LMIs requires convex optimisation algorithms and the size of problems which can be handled with current algorithms is quite limited.

H_∞ and LQG optimal control both suffer from the deficiency that the mapping from the weights used to tune them to a particular performance specification involving time-domain constraints is obscure. This means that an additional optimisation search over the weights would be needed to evaluate controllability.

A more promising methodology is ℓ_1 optimal control. Using the minimisation of the ℓ_1 norm to optimise disturbance rejection was first introduced by Vidyasagar (1986). The extension of the ℓ_1 problem to discrete MIMO systems was given in Dahleh and Pearson (1987b). This method which is presented, in depth, in Dahleh and Diaz-Bobillo (1995) has a number of particular characteristics which recommend it for consideration for controllability analysis.

- It deals directly with time-domain constraints and disturbances with time-domain bounds and, hence, has a natural fit with the typical performance requirements which might be specified.
- It optimises the achievable performance over the set of linear time invariant (LTI) controllers (square or non-square), which is a broad and realisable set.
- It incorporates most of the fundamental limitations on controllability, discussed in section 2.1, directly in the optimisation problem.
- It requires the solution of a *linear* program, rather than a general convex program, and is therefore potentially applicable on realistically sized problems.

Therefore the framework of this problem satisfies many of the requirements, mentioned in 2.4.1, for a new linear controllability technique. This ℓ_1 technique provides a very strong unambiguous result. It answers the question: Is there any linear time invariant controller which can keep the constrained variables within their bounds for the worst time-bounded disturbance? This method and the techniques used to solve it have been used as a basis for the linear controllability analysis technique presented in this chapter.

3.2 Achievable Closed Loop Transfer Functions

The closed-loop system is defined by the relationships below, where P is the generalised plant, K is the controller, w are the disturbances, u are the control inputs, z are the regulated outputs and y are the measured outputs.

$$
\begin{aligned}
z &= P_{11}w + P_{12}u \\
y &= P_{21}w + P_{22}u \\
u &= Ky
\end{aligned}
$$

We take P and K to be linear time invariant transfer functions. z may contain any process variable including elements of u and y. y may include any of the disturbances, w, to incorporate feedforward. w may include measurement noise. The number of elements in each signal vector ($n_z,\ n_w,\ n_u,\ n_y$) are not required to satisfy any particular relationship. The generalised regulator problem is to choose K so as to optimise some property of the closed loop transfer function (mapping) from w to z while maintaining internal stability.

The mapping from w to z is given by

$$\Phi = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}. \tag{3.2}$$

The above is a special case of a more general description of feedback interconnections known as a lower Linear Fractional Transformation (LFT), ie, the mapping from w to z for the block diagram shown in Figure 3.1.

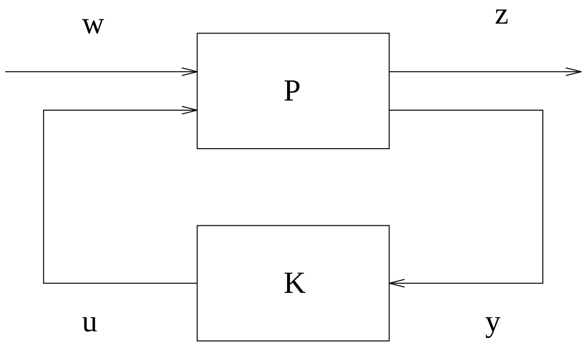

Figure 3.1: Block diagram of lower linear fractional transformation (LFT)

Ideally, the parametrisation of the set of stabilising controllers K would furnish the optimisation problem with the set of *all* stabilising controllers, so that all achievable closed-loop maps were provided. The parametrisation presented is often referred to as the Youla parametrisation (Youla *et al.*, 1976) and is linear fractional in a free stable parameter Q. If this parameter is chosen to belong to the ℓ_1 vector-space, ie, has a bounded ℓ_1-norm, then all LTI stabilising controllers are parametrised and all closed-loop maps are affine in Q. This enables the problem to be formulated as a linear program.

3.2.1 Parametrisation of LTI Stabilising Controllers

The controller is parametrised using the Youla parametrisation (Youla *et al.*, 1976), which parametrises all stabilising controllers in the stable parameter Q. This feedback controller will mean that the closed-loop system is internally stabilised. For this approach, which is based on double coprime factorisation, all stabilising controllers for the plant P_{22} are given by

$$K = (Y - MQ)(X - NQ)^{-1} = (\tilde{X} - Q\tilde{N})^{-1}(\tilde{Y} - Q\tilde{M}) \tag{3.3}$$

where Q is stable and N, M, $\tilde{N}$, $\tilde{M}$ are given by the right and left coprime factorisations

$$P_{22} = NM^{-1} = \tilde{M}^{-1}\tilde{N} \tag{3.4}$$

and stable X,Y,$\tilde{X}$,$\tilde{Y}$ satisfy

$$\begin{pmatrix} \tilde{X} & -\tilde{Y} \\ -\tilde{N} & \tilde{M} \end{pmatrix} \begin{pmatrix} M & Y \\ N & X \end{pmatrix} = I \tag{3.5}$$

If Q is chosen as $Q \in \ell_1$, i.e., has a finite ℓ_1 norm given by

$$\|Q\|_1 = \max_{1 \le i \le m} \sum_{j=1}^{n} \sum_{t=0}^{\infty} |q_{ij}(t)| \le \infty,$$

then all stabilising LTI controllers, in the ℓ_∞ sense, are parametrised by (3.3). If a controller is stabilising in the ℓ_∞ sense it means that the ℓ_1 norm of the closed loop system is finite.

When this controller parametrisation(3.3) is substituted into (3.2) the closed-loop map becomes affine in the free parameter Q.

$$\Phi = H - UQV \tag{3.6}$$

where

$$
\begin{aligned}
H &= P_{11} + P_{12} Y \tilde{M} P_{21}, \qquad (3.7)\\
U &= P_{12} M,\\
V &= \tilde{M} P_{21}.
\end{aligned}
$$

This controller parametrisation allows feedforward control to be incorporated in the problem, as well as feedback. This is done in a very straight forward manner. Simply augment the measured variables with the disturbance variables so that

$$\tilde{y} = \begin{pmatrix} y \\ w \end{pmatrix} = \tilde{P}_{21} w + \tilde{P}_{22} u \qquad (3.8)$$

where

$$\tilde{P}_{21} = \begin{pmatrix} P_{21} \\ I \end{pmatrix} \text{ and } \tilde{P}_{22} = \begin{pmatrix} P_{22} \\ 0 \end{pmatrix} \qquad (3.9)$$

and calculate your H, U and V, using (3.7), for these new $\tilde{P}_{21}$ and $\tilde{P}_{22}$.

3.3 Feasibility Constraints

There are two kinds of constraints for a given closed-loop map Φ, feasibility constraints and performance constraints. The former ensure that any selected Φ does indeed satisfy (3.6)

$$\Phi = H - UQV$$

for some stable Q. The latter represent the performance objectives and are discussed later in the chapter.

If P and K are expressed as discrete time LTI systems and Φ is a discrete time impulse response matrix, then it is possible to express these feasibility constraints as linear equality constraints. This chapter presents two such techniques, interpolation conditions and Q-approximation. The first method involves the use of a set of constraints on Φ, called the interpolation conditions, that ensure that the closed-loop map

Φ can be written in the form of (3.6) for some $Q \in \ell_1$. The second involves approximating Q as a finite length discrete impulse response. Both techniques provide linear constraints defining the achievable Φ.

3.3.1 Interpolation Conditions

Interpolation conditions provide necessary and sufficient conditions on the R in the closed-loop map

$$\Phi = H - R$$

such that it can be written as UQV for some stable Q. These conditions represent limitations in achieving the performance objectives. For example, if Q_d is chosen to give some desired value of Φ, ie, $H - UQ_dV =: \Phi_d$, then for any arbitrarily chosen Φ_d to be achievable Q must invert both U and V. This might not be possible due to either or both of U and V having RHP zeros, ie, not having stable inverses, or Q not having sufficient degrees of freedom, ie, a bad rank or nonsquare problem, thus limiting the achievable Φ. These limitations are captured by what are known as the zero and rank interpolation conditions respectively. The first set of conditions capture fundamental process limitations, such as RHP zeros or time delays. The second set depend on the objective function, since the selection of the exogenous disturbance w and the regulated output z will define whether the problem is one-block (good rank) or multi-block (bad rank).

These conditions can be expressed in two ways as presented in Dahleh and Diaz-Bobillo (1995), either as algebraic conditions on $R(\lambda)$ (Theorem 3.3.1) or conditions on the left and right null-spaces of the operator R (Theorem 3.3.2). Note that these theorems are given in terms of the discrete operator $\lambda = z^{-1}$, in which case the RHP zeros are given by the zeros λ_0 in the unit disc $\lambda_0 \in \mathcal{D}$.

Theorem 3.3.1 *Assume that $\Lambda_{UV} \subset \mathcal{D}$. Given $R \in \ell_1^{n_z \times n_w}$, there exists a $Q \in \ell_1^{n_u \times n_y}$ such that $R = UQV$ if and only if $\forall \lambda_0 \in \Lambda_{UV} \subset \mathcal{D}$ the following conditions are satisfied:*

$$i) \quad (\alpha_i R \beta_j)^{(k)}(\lambda_0) = 0 \quad for \quad \begin{cases} i = 1, ..., n_u \\ j = 1, ..., n_y \\ k = 0, ..., \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0) - 1 \end{cases} \tag{3.10}$$

$$ii) \quad \begin{cases} (\alpha_i R)(\lambda) \equiv 0 \quad for \quad i = n_u + 1, ..., n_z \\ (R\beta_j)(\lambda) \equiv 0 \quad for \quad j = n_y + 1, ..., n_w \end{cases} \tag{3.11}$$

where $\alpha_i(\lambda) = (L_U^{-1})_i(\lambda) \quad i = 1, 2, ...n_z$, $\beta_j(\lambda) = (R_V^{-1})^j(\lambda) \quad j = 1, 2, ...n_w$ for the Smith-McMillan decompositions $U = L_U M_U R_U$ and $V = L_V M_V R_V$, $\mathcal{D}$ =open unit disc,$\bar{\mathcal{D}}$ =closed unit disc, Λ_{UV} =set of zeros of U and V in $\bar{\mathcal{D}}$, σ_{U_i} =sequence of structural indices corresponding to U, ie,$\sigma_{U_i}(\lambda_0)$=multiplicity of the zero λ_0 as a root of the numerator of the ith diagonal term of M_U, σ_{V_j} =sequence of structural indices corresponding to V.

The conditions in i) are the zero interpolation conditions, whilst the conditions in ii) are the rank interpolation conditions. Note however that there are an infinite number of constraints involved in implementing ii), since

$$\begin{aligned} (\alpha_i R)(\lambda) &= \sum_{t=0}^{\infty} (\alpha_i R)(t)\lambda^t \equiv 0 \Rightarrow (\alpha_i R)(t) = \underline{0} \;\; \forall t >= 0 \\ (R\beta_j)(\lambda) &= \sum_{t=0}^{\infty} (R\beta_j)(t)\lambda^t \equiv 0 \Rightarrow (R\beta_j)(t) = \underline{0} \;\; \forall t >= 0 \end{aligned} \tag{3.12}$$

Therefore for the constraints to be implementable the problem must have no rank conditions. This is so for one-block (good rank or square) problems where $n_w = n_y$ and $n_z = n_u$, but not for multiblock (bad rank or nonsquare) problems where $n_w > n_y$ and $n_z > n_u$. Therefore, if the problem is multi-block, it is embedded in a one-block problem which has only zero interpolation conditions. In this work the delay augmentation (DA) algorithm is used, which involves U and V being augmented by N pure delays, so that the problem is one-block. As N increases the zero interpolation conditions of the DA problem approximate the original problem increasingly closely. This is discussed in more detail in Appendix A.1.

The Smith-McMillan decomposition is used in theorem (3.3.1), since it provides a characterisation of the zero and pole structure of a rational matrix, however ,in order to avoid the computational difficulties of this decomposition (numerically sensitive), an alternative expression of the zero interpolation conditions can be given using null chains. Hence, equivalently, the conditions, for a one-block problem (good rank), can be stated as follows:

Theorem 3.3.2 *Given a one-block problem, the zero interpolation conditions become:* $\forall\lambda_0 \in \Lambda_{UV} \subset \mathcal{D}$ *the following conditions must be satisfied:*

$$(y^i_{\lambda_0} R x^j_{\lambda_0})^{(k)}(\lambda_0) = 0 \quad for \quad \begin{cases} i = 1, ..., n_u \\ j = 1, ..., n_y \\ k = 0, ..., \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0) - 1 \end{cases} \tag{3.13}$$

$$where \;\; y^i_{\lambda_0}(\lambda) = \sum_{k=0}^{\sigma_{U_i}(\lambda_0)-1} (\lambda - \lambda_0)^k (y^i_{k+1})^T \;\; and \;\; x^j_{\lambda_0}(\lambda) = \sum_{k=0}^{\sigma_{V_j}(\lambda_0)-1} (\lambda - \lambda_0)^k x^j_{k+1}$$

for y^i *=elements of the extended set of left null chains of* U*,* x^j *=elements of the extended set of right null chains of* V*.*

Therefore the interpolation conditions can be given by a finite number of zero interpolation conditions of the kind shown in theorem (3.3.2) for a one-block problem. If the problem is multiblock then it will be solved, using the DA algorithm, as a series of one-block problems converging to the original problem.

Substituting $R = H - \Phi$ into equation (3.13) gives

$$(y^i_{\lambda_0}(H - \Phi) x^j_{\lambda_0})^{(k)}(\lambda_0) = 0 \quad for \quad \begin{cases} i = 1, ..., n_u \\ j = 1, ..., n_y \\ k = 0, ..., \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0) - 1 \end{cases}$$

which allow the elements of any feasible Φ to be expressed as follows

$$\sum_{p=1}^{n_z} \sum_{q=1}^{n_w} \sum_{l=0}^{\infty} a_{pq}^{ij,\lambda_0,k}(l) \phi_{pq}(l) = b^{ij,\lambda_0,k} \tag{3.14}$$

where

$$b^{ij,\lambda_0,k} = \sum_{p=1}^{n_z}\sum_{q=1}^{n_w}\sum_{l=0}^{\infty} h_{pq}(l). \tag{3.15}$$

and

$$a_{pq}^{ij,\lambda_0,k}(l) = \left[\sum_{t=0}^{\infty}\sum_{s=0}^{\infty}(y_{\lambda_0}^i)^p(t-l-s)(x_{\lambda_0}^j)_q(s)(\lambda^t)^{(k)}\right]_{\lambda=\lambda_0} \tag{3.16}$$

where $(x_{\lambda_0}^j)_q(k)$ and $(y_{\lambda_0}^i)^p(k)$ are the kth discrete elements of the qth component and pth component of the column vector $x_{\lambda_0^j}$ and the row vector $y_{\lambda_0}^i$ respectively.

These elements of $a_{pq}^{ij,\lambda_0,k}(l)$, $b^{ij,\lambda_0,k}$ and $\phi_{pq}(l)$ can be stacked and arranged into a constant real matrix A_{zero}, a constant real vector b_{zero} and a variable vector ϕ respectively so that the zero interpolation conditions expressed by equation (3.14) can be stated as

$$A_{zero}\phi = b_{zero} \tag{3.17}$$

This provides a finite number, c_z, of constraints,

$$c_z = \sum_{\lambda_0 \in \Lambda_{UV}} \sum_{i=1}^{n_u}\sum_{j=1}^{n_y} \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0), \tag{3.18}$$

however ϕ is infinite dimensional therefore this introduces an infinite number of variables. The method used to get a finite dimensional problem is to truncate ϕ for each z_i and w_j to a finite length N_{ij}. The elements of A_{zero} associated with ϕ beyond this horizon are removed from the problem. This implies that any ϕ satisfying the truncated constraints also satisfies the infinite constraints for $\phi_{ij}(k) = 0\ \forall k \geq N_{ij}$. This technique (Dahleh and Diaz-Bobillo, 1995) is given in more detail in Appendix A.2.

The effect of the choice of N and N_{ij} will be discussed later when the problem definition is complete.

3.3.2 Q-Approximation

There is another technique, which avoids calculating these interpolation conditions at all. For this technique the discrete impulse response of Q is approximated as

$$Q = \sum_{p=0}^{L-1} q(p)z^{-p} \tag{3.19}$$

(Boyd *et al.*, 1988) with finite length L. The assumption that Q is a finite impulse response of length L, means that this technique is an approximation and it is necessary to select L large enough to include the optimal Q. Using this approximation means that the discrete elements of Φ can be expressed as follows

$$\begin{aligned}
\Phi &= H - UQV \\
\Rightarrow \phi(k) &= h(k) - (\sum_{r=0}^{k}\sum_{p=0}^{r} u(r-p)q(p)v(k-r)) \\
&= h(k) - (\sum_{r=0}^{k}\sum_{p=0}^{r} u(r-p)\sum_{m=1}^{n_y} q^m(p)v_m(k-r)) \\
&= h(k) - (\sum_{r=0}^{k}\sum_{p=0}^{r}\sum_{n=1}^{n_u}\sum_{m=1}^{n_y} u^n(r-p)v_m(k-r)q_{nm}(p))
\end{aligned}$$

which can be rearranged to give the elements $\phi_{ij}(k)$ of ϕ as follows

$$\phi_{ij}(k) + \left[\sum_{p=1}^{\min(k,L-1)}\sum_{n=1}^{n_u}\sum_{m=1}^{n_y} c_{nm}^{ij(k)}(p)q_{nm}(p)\right] = h_{ij}(k) \quad \text{for} \quad \begin{cases} i = 1, ..., n_u \\ j = 1, ..., n_y \\ k = 0, ..., \infty \end{cases} \tag{3.20}$$

where

$$c_{nm}^{ij(k)}(p) = \sum_{r=p}^{k} u_{in}(r-p)v_{mj}(k-r) \tag{3.21}$$

Much as in the zero interpolation case these elements of $c_{nm}^{ij(k)}(p)$, $h_{ij}(k)$, $\phi_{ij}(k)$ and $q_{nm}(p)$ can be stacked and arranged into a real constant matrix A_q, a real constant vector h and variable vectors ϕ and q respectively . In this way expression (3.20) can be stated as

$$[I \quad | \quad A_q]\begin{bmatrix} \phi \\ \hline q \end{bmatrix} = h \tag{3.22}$$

Note that each SISO impulse response $q_{nm}(k)$ could be assigned a distinct length L_{nm}, but this would normally introduce an unhelpful increase in complexity.

As with the zero interpolation constraints ϕ is truncated and the elements of ϕ_{ij} beyond N_{ij} ignored. In this case this implies nothing about the values of $\phi_{ij}(k)$ $k \geq N_{ij}$.

3.4 Disturbance Description

In typical process applications the disturbances will not be precisely characterised. However the following questions can usually be answered for the key disturbances. What range of values does the disturbance take? How quickly can the value change? Do changes occur frequently? Based on answers to these questions, an adequately precise disturbance description can be developed.

Very slowly changing disturbances should be treated as steady-state. Disturbances which change significantly over the process transient response time can be modelled as instantaneous changes filtered through a transfer function W_1. If the maximum rate of change is 10% of maximum variation per sample, a finite impulse response (FIR) filter W_1,

$$W_1 = 0.1 \sum_{i=0}^{9} z^{-i}$$

would convert a step change between the limits to a ramp at the maximum rate, thereby excluding the unrealistic disturbances. Such filters can be included in the process model as shown in figure 3.2.

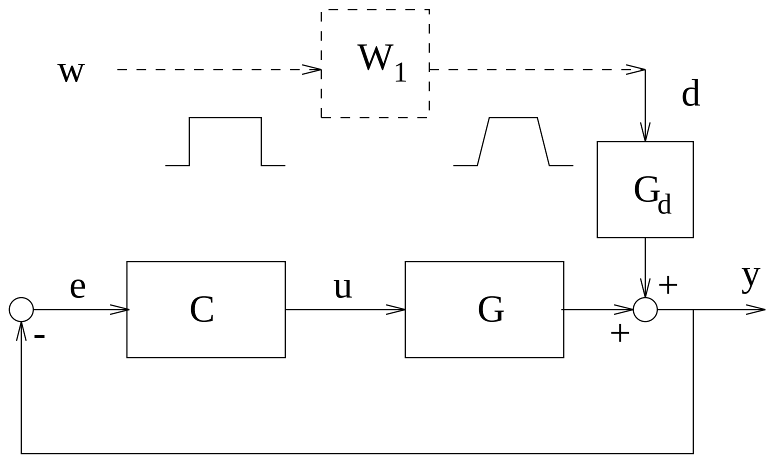

Figure 3.2: Block diagram for the disturbance filter augmented system

If changes occur infrequently, compared to the process response time, then a step disturbance description is more appropriate than a persistently varying signal which is only constrained to be magnitude bounded.

Typical time-domain disturbance descriptions can therefore be represented in terms of three types of disturbance signals: persistently varying (persistent) disturbances, step disturbances and steady-state disturbances.

Given a finite impulse response description of ϕ we can compute the worst-case response for each of these signal types. For convenience (and without loss of generality) we consider all disturbances relative to a reference point corresponding to $w_{centre} = (w^h + w^l)/2$ and scaled so that a unit variation corresponds to $(w^h - w^l)/2$. This causes no loss in generality, since for linearised systems with asymmetric bounds on the disturbance the system can simply be rescaled about the reference point w_{centre} as for the case study in 5.2. The disturbance deviations about this point will be symmetric with a maximum deviation of unity (upper bound = -lower bound = 1) and the corresponding output deviations will also be symmetric.

If the problem involves measurement noise, then the noise should be represented as a disturbance in the disturbance vector w and this noise-disturbance added to the relevant measurement y in the system description. A suitable description of the noise should be selected from the possible disturbance descriptions of persistently varying, step or steady-state.

3.4.1 Steady-State Disturbances

The maximum output deviation on z_i in response to w_j will be given by

$$-\nu_{ij} \leq \sum_{k=0}^{\infty} \phi_{ij}(k) \leq \nu_{ij} \tag{3.23}$$

However we will only have access to ϕ_{ij} up to a finite horizon N_{ij}. With the interpolation constraints the later values of ϕ_{ij} have been set to zero and therefore the finite horizon sum equals the infinite horizon sum. With the Q-approximation the finite sum is not equal to the infinite sum. However, the steady-state gain of Q can be evaluated as a finite sum and the corresponding steady-state gain of ϕ computed given the steady-state values of H, U, and V. Hence the closed loop steady-state gains, $G_{zw}^{ss,cl}$, and ν_{ij}, can be computed using finite dimensional linear constraints in both cases.

3.4.2 Step Disturbances

We consider steps between any two levels within the lower and upper bounds on the disturbance value. The worst-case will be given by a step between the upper and lower bound or vice versa. The effect on output z_i of a step in disturbance w_j from the lower bound (-1) to the upper bound (1) would be,

$$z_{ij}(t) = -z_{ij}^{ss} + 2\left(\sum_{k=0}^{t} \phi_{ij}(k)\right) = -\sum_{k=0}^{\infty} \phi_{ij}(k) + 2\left(\sum_{k=0}^{t} \phi_{ij}(k)\right)$$

The maximum deviation is therefore given by

$$-\nu_{ij} \leq -\sum_{k=0}^{\infty} \phi_{ij}(k) + 2\left(\sum_{k=0}^{t} \phi_{ij}(k)\right) \leq \nu_{ij} \quad \text{for} \quad \begin{cases} t \geq 0 \\ i = 1, \ldots.., n_z \\ j = 1, \ldots.., n_w \end{cases} \tag{3.24}$$

The steady-state component can be computed as above. With the interpolation constraints the maximum deviation for $t < N_{ij}$ will be the maximum deviation for the infinite horizon. With the Q-approximation the peak step response calculated in this way will be a lower bound on the infinite horizon step response deviation as the later values of ϕ are unconstrained.

3.4.3 Persistent Disturbances

For persistently varying disturbances we have

$$\nu_{ij} = \sum_{k=0}^{\infty} |\phi_{ij}(k)|$$

The absolute value in the above expression requires a change of variable to give a suitable form for a LP. We define $\Phi = \Phi^{+} - \Phi^{-}$, where Φ^{+} and Φ^{-} have only nonnegative elements. We can then write

$$\nu_{ij} = \sum_{k=0}^{\infty} \phi^{-}{}_{ij}(k) + \phi^{+}{}_{ij}(k) \tag{3.25}$$

Replacing the sum to infinity with the sum to N_{ij} has the same implications as for the step disturbances.

3.4.4 Combining Disturbances

For each disturbance type we can calculate ν_{ij}. For a purely worst-case analysis we get the worst deviation on z_i as

$$\nu_i = \sum_{j=1}^{n_w} \nu_{ij} \tag{3.26}$$

If such an analysis is felt to be excessively conservative then we may define ν_i based on the maximum deviation over specified subsets of disturbances, e.g. any n step disturbances combined with all the steady-state and persistent disturbances. The constraints for each allowable combination could be appended to the LP.

The values of ν_i couple the disturbance description with the overall performance requirements.

3.5 Performance Requirements

For the linear controllability problem in (3.1)

$$\min_{K,u_0} J(K,u_0) \;\; s.t. \;\; c(K,u_0,w) \leq 0 \; \forall \; w \in W$$

conditions which ensure that the controller, K, is selected from the set of all stabilising LTI controllers have been discussed in section 3.3. These conditions can be represented by a set of finite linear constraints on the discrete impulse response of Φ (the closed loop map from the disturbances w to the regulated outputs z). Therefore this problem could be restated as

$$\min_{\Phi,u_0} J(\Phi,u_0) \;\; s.t. \;\; \begin{cases} c(\Phi,u_0,w) \leq 0 \; \forall \; w \in W \\ \Phi(K) \text{ where K=LTI stabilising controller} \end{cases}$$

To solve this as an LP both the objective, J, and the performance constraints, c, should be given as linear functions on Φ and u_0.

The constraints and objective should be formulated to capture the performance requirements for a specific problem and plant. Typical process constraints such as magnitude bounds on process variables, ie, a certain temperature must be kept between an upper and lower bound, can be captured in a straightforward manner. This chapter will discuss formulation of the LP to satisfy a range of objectives and performance constraints.

We assume that the model used has been linearised about a steady-state given by z_{lin}, y_{lin}, w_{lin}, u_{lin} and that all constraint limits have been expressed relative to this steady-state. As mentioned in section 3.4 all disturbances are considered relative to a reference point w_{centre} and scaled so that disturbance deviations about this point will be symmetric with a maximum deviation of 1. Also the corresponding output deviations will be symmetric about a value of z, corresponding to w_{centre} and the operating point u_0, which is given by

$$z^{ref} = G_{zw}^{ss,ol}(w_{centre} - w_{lin}) + G_{zu}^{ss,ol}(u_0 - u_{lin}) + z_{lin} \tag{3.27}$$

where $G^{ss,ol}$ denotes the open-loop steady-state gains and $G^{ss,cl}$ would denote the closed loop steady-state gains. Similarly the value of y corresponding to w_{centre} and u_0 is given by

$$y^{ref} = G_{yw}^{ss,ol}(w_{centre} - w_{lin}) + G_{yu}^{ss,ol}(u_0 - u_{lin}) + y_{lin} \tag{3.28}$$

where the controller is given by $u = u_0 + K(s)(y - y^{ref})$.

The most basic performance requirement is to ensure that none of the variables violate their bounds. This can be enforced by including the constrained variable in z and enforcing

$$z_i^l \leq z_i \leq z_i^h \quad \forall w \in W \; i = 1, n_z \tag{3.29}$$

Hence, to enforce input constraints, simply include the inputs of interest from u into z. For example, to include the pth control input u_p, alter the original system description

$$\begin{aligned} z &= P_{11}w + P_{12}u \\ y &= P_{21}w + P_{22}u \end{aligned}$$

to the following

$$\begin{pmatrix} z \\ u_p \end{pmatrix} = \begin{pmatrix} P_{11} \\ 0 \end{pmatrix} w + \begin{pmatrix} P_{12} \\ e_p^T \end{pmatrix} u$$

$$y = P_{21}w + P_{22}u$$

where $e_p = p$th column of $I_{n_u \times n_u}$.

The maximum deviation in z_i due to the disturbances is given by ν_i,

$$\nu_i = \sum_{j=1}^{n_w} \nu_{ij} \tag{3.30}$$

where ν_{ij}, the effect due to the jth disturbance w_j, is calculated in the LP according to the set of linear constraints appropriate to the description of w_j, i.e., whether it is steady-state (see 3.4.1), a step (see 3.4.2) or persistent (see 3.4.3). As mentioned previously this maximum deviation is applied symmetrically about the reference point z_{ref} in (3.27) therefore (3.29) can be rewritten as

$$z_i^l + \nu_i \leq z_i^{ref} \leq z_i^h - \nu_i \quad i = 1, n_z \tag{3.31}$$

Therefore the constraints $c(K, u_0, w) \leq 0$ would be given by

$$\left.\begin{array}{rcl} z_i^l + \nu_i - z_i^{ref} & \leq & 0 \\ z_i^{ref} - z_i^h + \nu_i & \leq & 0 \end{array}\right\} \forall i$$

Rather than imposing the performance requirements directly we may wish to determine the fraction of the disturbance deviations for which the performance is achievable. This can only be expressed as a linear program if z^{ref} is independent of K and u_0, i.e., for a fixed operating point u_0. In this case we can scale the deviations z_i by the minimum distance to a constraint $(\min(z_i^h - z_i^{ref}, z_i^{ref} - z_i^l))$ so that a deviation less than 1 corresponds to constraint feasibility. We can then make the objective of the LP.

$$\nu_o = \min_K \max_{i=1,n_z} \nu_i \qquad (J(K) = \max_{i=1,n_z} \nu_i) \tag{3.32}$$

If the disturbances are scaled down relative to w_{centre} by a factor $\frac{1}{\nu_o}$ then the performance constraints will be unviolated. Therefore $\frac{1}{\nu_o}$ gives the fraction of the disturbance set which can be tolerated. The use of the "disturbance fraction" as a controllability measure is discussed more generally in Walsh and Perkins (1996).

If we have a reasonable basis for penalising the maximum deviation on each objective z_i, then it could be appropriate to minimise a weighted sum of the deviations ν_i, rather than the maximum deviation. The weighted deviations could be minimised subject to feasibility constraints. This type of objective is highly flexible. Its key limitation is the difficulty of defining appropriate weights for the deviations.

In sections 3.5.1 and 3.5.2 two specific problems, that can be solved using the formulation of (3.1), are discussed. The first is the ℓ_1 optimal control problem, which is an existing technique first introduced in Vidyasagar (1986) and on which much work, by a variety of people, has been carried out. The second is the OLDE problem, which has been developed as part of the work for this thesis and which can be used to assess the economic performance of a linear dynamic system.

3.5.1 The ℓ_1 Performance Problem

If all the disturbances are persistent, then ν_0, described in (3.32), is the ℓ_1 norm of the mapping from w to z and the problem being solved is an ℓ_1 optimal control problem. The ℓ_1 optimal control problem involves the minimisation of the ℓ_1 norm of an objective function over the set of all stabilising LTI controllers. This problem has been well researched and the work discussed in the following is based mostly on work presented in papers by Vidyasagar (1986) and Dahleh and Pearson (1986; 1987b; 1987a) and a book by Dahleh and Diaz-Bobillo (1995).

The ℓ_1 optimal control problem selects a stabilising LTI controller, to minimise the maximum deviation in the regulated outputs z, over all time, due to unity magnitude bounded disturbances. This allows us to address the following controllability question: Given a nominal linear model of a process and a set of disturbances, which are bounded over time, is there any stabilising LTI controller for which the maximum deviation of the process variables over time is acceptable? Using ℓ_1 optimal control to solve an ℓ_1 performance optimisation problem as a process controllability analysis tool is discussed in Chenery and Walsh (1996).

The problem to be solved is

$$\inf_{K stab.} \sup_{\|w\|_\infty <= 1} \|z\|_\infty = \inf_{K stab.} \sup_{\|w\|_\infty <= 1} \|\Phi w\|_\infty \tag{3.33}$$

where the infinity norm of z is the maximum deviation in z over all (discrete) time,

$$\|z\|_\infty = \sup_k \max_i |z_i(k)|. \tag{3.34}$$

and similarly for w. The set of disturbances is very general in that they are unknown, but magnitude bounded by 1, i.e., $\|w\|_\infty \leq 1 \quad \Rightarrow w \in \ell_\infty$. The induced norm for operators on ℓ_∞ is the ℓ_1-norm

$$\|\Phi\|_1 = \sup_{\|w\|_\infty \leq 1} \|\Phi w\|_\infty. \tag{3.35}$$

which is computed using,

$$\|\Phi\|_1 = \max_{1 \leq i \leq n_z} \sum_{j=1}^{n_w} \sum_{t=0}^{\infty} |\phi_{ij}(t)| \tag{3.36}$$

Substituting (3.35) into (3.33) gives

$$\inf_{K stab.} (\sup_{\|w\|_\infty \leq 1} \|\Phi w\|_\infty) = \inf_{K stab.} \|\Phi\|_1 \tag{3.37}$$

the ℓ_1 optimal control problem.

The description of all disturbances as unknown, but magnitude bounded, ie, $w \in \ell_\infty$, allows the inclusion of persistent disturbances which could not be described as finite energy signals, ie, $w \in \ell_2$. However this is a worst-case design method, since the disturbance can take any form as long as its magnitude is bounded, which means that the measure it provides might be conservative.

The ℓ_1 performance problem should be set up as follows,

$$\nu_o = \inf_{K stab.} \|\Phi\|_1. \tag{3.38}$$

The objective function $\|\Phi\|_1$ is nonlinear due to the absolute norm in (3.36), to avoid this nonlinearity a variable change is made. Put $\Phi = \Phi^+ - \Phi^-$ where Φ^+ and Φ^-

are sequences of maps with non-negative entries. Using this it can be shown that the ℓ_1-norm of Φ is equivalent to,

$$\max_i \sum_{j=1}^{n_w} \sum_{t=0}^{\infty} (\phi_{ij}^{+}(t) + \phi_{ij}^{-}(t)) = \nu. \tag{3.39}$$

Therefore the ℓ_1 minimisation problem (3.38) can be stated as follows,

$$\nu_o = \inf_{\nu, \Phi^+, \Phi^-} \nu$$

subject to

$$\mathcal{A}_{\ell_1}(\Phi^+ + \Phi^-) \leq \mathbf{1}\nu \tag{3.40}$$

$$\Phi = \Phi^+ - \Phi^- \text{ is feasible.}$$

where the operator $\mathcal{A}_{\ell_1}$ is defined such that

$$(\mathcal{A}_{\ell_1}\Phi)_i = \sum_{j=1}^{n_w} \sum_{t=0}^{\infty} \phi_{ij}(t) \tag{3.41}$$

for $i = 1, \ldots\ldots, n_z$ and $\mathbf{1}$ is a vector all of ones $\in \mathcal{R}^{n_w}$.

3.5.2 The Optimal Linear Dynamic Economic (OLDE) Problem

In general, process design involves a steady state optimisation, which minimises the appropriate objective function by selecting a set of steady state operating values subject to a set of equality and inequality constraints. Often this optimisation will result in plant operation on these operational inequality constraints. However, since some process disturbances will cause these constraints to be violated, it will not be possible to operate the plant on them. To avoid this violation during operation, the steady state operating point will have to be moved sufficiently far away from the optimum and into the feasible region. This backing off from the optimal steady-state operating point reduces economic performance.

Ideally controllability analysis should consider, not only feasibility of a design, but also this economic performance. Therefore it would be desirable to incorporate operating economics into any controllability analysis technique as a measure of achievable performance.

If we linearise around the steady-state optimum for $\bar{w} = w_{centre}$ ($\bar{w}$ is the expected disturbance level) and the active set of constraints is assumed to remain constant over the disturbance set, then the economic penalty can be estimated as (Narraway *et al.*, 1991)

$$J = J_o - \sum_{i=1}^{n} \lambda_i \delta_i .$$

J is the objective function at the optimal operating point, J_o is the steady state optimum, λ_i is the Lagrange multiplier of the constraint on z_i and δ_i is the back off required to compensate for the maximum deviation of the ith constraint due to disturbances. Therefore let $|\delta_i| = \nu_i$, where the ν_i's satisfy the appropriate linear constraints presented in section 3.4. Hence the objective of the LP becomes

$$\nu_o = \min_{\nu_i, \tilde{\phi}^+_{ij}, \tilde{\phi}^-_{ij}} \sum_{i=1}^{n_z} \lambda_i \nu_i$$

Alternatively the linearised steady-state constraints and objective function can be added to the LP formulation. This allows changes in the active set of constraints to be accommodated, subject to the limitations of the linear steady-state model used. If we have reasonable information about the economics of the process then we can determine the optimal expected value of this objective over the disturbance set. By virtue of closed-loop linearity, the expected value of the objective is simply the objective for the expected value of the disturbance, $\bar{w}$, i.e.,$E(o(w)) = o(E(w)) = o(\bar{w})$.

Therefore if the objective function, $o(w, u)$, for the expected value of the disturbance is expressed as

$$o(\bar{w}) \quad | \quad u = u_0 + K(s)(y - y^{ref}) \tag{3.42}$$

$$o(\bar{w}) \quad = \quad o^{ref} + G^{ss,cl}_{ow}(\bar{w} - w_{centre}) \tag{3.43}$$

where

$$o^{ref} = G^{ss,ol}_{ow}(w_{centre} - w_{lin}) + G^{ss,ol}_{ou}(u_0 - u_{lin}) + o(w_{lin}, u_{lin}) \tag{3.44}$$

then our optimisation becomes

$$\min_{K, u_0} o^{ref} + G^{ss,cl}_{ow}(\bar{w} - w_{centre}) \tag{3.45}$$

Note that if o is included in z then $G_{ow}^{ss,cl}$ can be constructed as discussed previously. Alternatively, if o is not in z and all inputs appear in z, then $G_{ow}^{ss,cl}$ can obtained as $G_{ou}^{ss,ol} G_{uw}^{ss,cl} + G_{ow}^{ss,ol}$.

Hence the optimisation problem (3.1) becomes

$$\min_{K,u_0} J(K, u_0) = \min_{K,u_0} o^{ref} + G_{ow}^{ss,cl}(\bar{w} - w_{centre})$$

where u_0 is chosen to satisfy the constraints $c(K, u_0, w) \leq 0$ described in (3.31)

$$\left.\begin{array}{rcl} z_i^l + \nu_i - (G_{zw}^{ss,cl}(w_{centre} - w_{lin}) + G_{zu}^{ss,ol}(u_0 - u_{lin}) + z_{lin})_i & \leq & 0 \\ (G_{zw}^{ss,cl}(w_{centre} - w_{lin}) + G_{zu}^{ss,ol}(u_0 - u_{lin}) + z_{lin})_i - z_i^h + \nu_i & \leq & 0 \end{array}\right\} \forall i$$

and K is chosen to minimise ν_i. This ensures that the operating point backs off sufficiently to avoid violation of the constraints by deviations in z_i's $\forall w \in W,\ i = 1, n_z$.

Young *et al.* (1996) present a conceptually similar approach to evaluating operating economics based on the Q-approximation. Their approach diverges significantly in detail. They do not consider the use of interpolation constraints and consider only step disturbances and nominal economics. Their disturbance set is defined by

$$\Gamma = \{(d_b, \Delta d) | d = d_b + \Delta d; d^- \leq d \leq d^+\}$$

for which the base value d_b, the magnitude and direction of the step Δd are uncertain. They reduce the set of disturbances to be considered when testing for feasibility to just the critical (vertex) points $(d^-, d^+ - d^-)$ and $(d^+, -(d^+ - d^-))$. For a large number of disturbances this set would still become large, 2^{n_w}. By choosing the reference disturbance as d_{centre}, we avoid the need to consider the above critical points individually. They allow convex objective functions, which means that their problem size is limited by the capacity of convex programming algorithms, rather than linear programming algorithms which can handle larger problems. They apply their method successfully to a simple problem. Despite the differences noted, their work is very much in the same spirit as the OLDE work.

3.6 Properties of the LP Formulations

We have shown that many performance requirements of practical interest can be formulated in terms of a LP, which can be solved to determine the optimal controller K and, if appropriate, the optimal operating point u_o. The LPs involve a number of approximations, so it is important to consider how the results of the LPs relate to the solution of the original infinite horizon discrete time problem.

In the interpolation constraint method, the finite horizon response is an upper bound on the infinite horizon response, due to the additional requirement that the zero impulse response beyond the horizon is achievable. If delay augmentation is used with finite N then the resulting solution is a lower bound on the finite horizon problem. For a single block ℓ_1 problem, it is possible to calculate the values for N_{ij} such that the constraints $\phi_{ij}(k) = 0 \ \forall k > N_{ij}$ are inactive (Dahleh and Diaz-Bobillo, 1995) and the finite horizon optimum becomes equal to the infinite horizon optimum. For a multi-block ℓ_1 problem this calculation can be carried out for each value of N and the finite horizons N_{ij} can be chosen accordingly. The solutions as N increases will then provide a sequence of lower bounds which converge to the infinite horizon optimum. For problems other than the ℓ_1 problem, the minimum value of N_{ij} needed to make the finite horizon optimum equal to the infinite horizon optimum is not known. The solutions generated are therefore lower bounds on an upper bound and hence simply approximations to the infinite horizon problem.

In the Q-approximation method, the optimal finite horizon response is a lower bound on the optimal infinite horizon response (with the same parameterisation of Q) since the finite horizon problem disregards potential deviations beyond the horizon. If L, the length of the finite impulse response of Q, is less than any horizon $N_{ij} + 1$ the resulting solution is an upper bound on the finite horizon problem, with arbitrary Q, as elements of Q affecting the finite horizon response are restricted to a value of zero. For $L \geq N_{ij} + 1 \ \ \forall i, j$ the solution for the Q-parameterisation is a lower bound on the infinite horizon problem if the finite horizon objective and constraints are not affected by $q_{mn}(k)$ being zero for $k > L - 1$. This condition is not guaranteed for step or steady-state disturbances and the Q-approximation gives just an approximate result for these types of disturbances.

With both methods, if the solution for varying L or N takes an almost constant value as L or N vary over most of the horizon, then the behaviour beyond the end of the horizon can be presumed not to be significant and the result can be taken as a true estimate of the achievable performance. This convergence can be tested by reconstructing the controller K and calculating the infinite horizon response to within a desired precision (Dahleh and Diaz-Bobillo, 1995) to give an upper bound on the infinite horizon solution.

In Ajbar *et al.* (1995) the interpolation constraint approach is criticised for giving only an approximate solution to the ℓ_1 problem, in contrast to the Q-approximation. However, both methods can be used to construct a sequence of lower bounds to the infinite horizon optimum or to construct upper bounds based on a specific solution. Also, both methods exhibit the qualitative property that consistent results as N or L increase indicates convergence to the true optimum. At least for the ℓ_1 problem both methods seem to be equally well founded. If the problem is one-block then it is more efficient and exact to solve the zero interpolation condition problem.

There are also differences in the structure of the associated linear programs. The zero interpolation conditions generate

$$\sum_{\lambda_0 \in \Lambda_{UV}} \sum_{i=1}^{n_z} \sum_{j=1}^{n_w} \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0) \qquad (\sigma_{U_i}(\lambda_0), \sigma_{V_j}(\lambda_0) = \text{ structural indices of } U,\ V)$$

constraints and

$$\sum_{i=1}^{n_z} \sum_{j=1}^{n_w} N_{ij} + 1 \qquad (N_{ij} = \text{ FIR length of } \phi_{ij})$$

variables within the LP.

The Q-approximation method generates

$$\sum_{i=1}^{n_z} \sum_{j=1}^{n_w} N_{ij} + 1$$

constraints and

$$\left[\sum_{i=1}^{n_z} \sum_{j=1}^{n_w} N_{ij} + 1 \right] + \left[\sum_{n=1}^{n_u} \sum_{m=1}^{n_y} L_{nm} + 1 \right] \qquad (L_{nm} = \text{ FIR length of } q_{nm})$$

variables. On the whole, the number of constraints in the zero interpolation conditions will be less than the number of variables, in which case the interpolation condition will produce a smaller LP problem. However the Q-approximation problem, on the whole, will be more sparse, which produces an LP which is faster to solve. Also if the DA algorithm must be used to produce the one-block problem, then the number of constraints produced by the zero interpolation conditions will grow as the number of delays augmenting the problem are increased in the algorithm. Therefore which method is preferable, in terms of LP complexity, is highly problem specific.

The Q-approximation is by far the simpler and more straight forward of the two techniques to implement. There are examples of various problems solved by each technique in the literature, e.g., (Ajbar *et al.*, 1995; Young *et al.*, 1996; Swartz, 1994) for the Q-approximation and (Dahleh and Diaz-Bobillo, 1995) for the zero interpolation method. We solved the very simple SISO problem presented in example 1 of (Ajbar *et al.*, 1995) using both techniques. Several iterations of the Q-approximation technique were required to check if a good choice of L had been made. The interpolation constraint technique solved immediately and exactly since the problem is one-block. In other examples, including the case study in 5.2 we have successfully used interpolation constraints while the Q-approximation LP has not solved successfully.

It should be noted, that both approaches have a variant in which finite step responses replace the finite impulse response (see section 3.7). For step and steady-state disturbances this will give a sparser LP and may have better solution characteristics.

It is clear from the above discussion that there is an element of approximation in the LP formulation. We believe that provided convergence of estimates is obtained for varying L or N this approximation is of little consequence. The approximation of a nonlinear process by a linear model is of primary importance, followed by the approximation of the continuous time linear model by a discrete time model. Variable transformations should be used where appropriate to improve the quality of the linear model and sampling should be faster than both open-loop transients in response to disturbances and the fastest expected closed loop response time (for examples of this see Chapter 5). Some iteration might be necessary to achieve this.

3.7 Extensions to this Work by Others

Two main extensions to the analysis presented in this chapter have been made by others. One is the specialisation of the OLDE problem to steady state systems for improved efficiency, this gave rise to the optimal linear steady state economic (OLSE) problem, and the other is the use of the OLDE framework with extra constraints and objective to give an economic evaluation of quality.

The first technique was set up by P. Owen and is applied to the case study in Walsh *et al* (1997). This involves solving the same type of problem as for the OLDE method, but with the system described as

$$\Phi^{ss} = H^{ss} - U^{ss} Q^{ss} V^{ss}, \quad Q^{ss} < N_Q$$

where Φ^{ss}, H^{ss}, U^{ss}, Q^{ss} and V^{ss} are all steady state gains and N_Q is a large finite number. This is solved efficiently as an LP using a simplified steady-state version of the Q-parametrisation technique. The technique reveals whether, for the worst steady state disturbance, the system can fulfill the performance requirements at steady state. If the problem fails this analysis there is no point analysing its dynamic behaviour. This problem is computationally inexpensive in comparison to the OLDE problem and therefore is a useful steady state screening test.

The second technique was developed by an ERASMUS student O. Frank (1997) for his Diploma Thesis and attempts to assess the achievable economic cost with respect to quality. The technique is based on the framework of the OLDE problem and is briefly summarised in the following.

This technique attempts to assess the achievable economic cost with respect to quality. The quality cost function, J_i^q, is represented by a discontinuous function on a variable, x, which is the deviation in the product from specification.

$$J^q(x) = \begin{cases} 0 & |x| < \nu_s \\ C_1 & |x| \geq \nu_s \end{cases}$$

where ν_s is the value above which a deviation in the product means it is off-specification and C_1 is chosen to represent the cost of reprocessing the off-specification product. For a maximum deviation in output z_i of ν_i from the operating point z_i^{ref} the output is assumed to have a normal distribution described by

$$f_i(x) = \frac{4}{\nu_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{4(x - z_i^{ref})}{\nu_i}\right)^2\right)$$

Clearly there can be a different quality cost function J_i^q for each output. The estimated cost of quality for the ith output is then given by

$$E(\nu_i, z_i^{ref}) = \int_{z_i^{ref}-\nu_i}^{z_i^{ref}+\nu_i} J_i^q(x) f_i(x) dx$$

which is convex for typical distributions, cost functions and percentages of acceptable off-specification production. Due to the convexity of this function $E(\nu_i, z_i^{ref})$ can be estimated as E_i through a set of linear constraints as follows

$$E_i \geq E(\nu_i^k, z_i^{ref,l}) + \left(\frac{\partial E}{\partial \nu_i}\right)_{\nu_i=\nu_i^k} + \left(\frac{\partial E}{\partial \nu_i}\right)_{z_i^{ref}=z_i^{ref,l}} \begin{cases} \forall k = 1,..,n \\ \forall l = 1,..,m \end{cases}$$

where ν_i^k and $z_i^{ref,l}$ are a finite set of values for which E and its partial derivatives are calculated. At each point defined by ν_i^k and $z_i^{ref,l}$ a linear plane tangential to $E(\nu_i^k, z_i^{ref,l})$ is found and we know that $E(\nu_i, z_i^{ref})$ must lie on and above this linear plane for all other values of ν_i and z_i^{ref}. If $\Delta\nu_i = \nu_i^k - \nu_i^{k-1}$ and $\Delta z_i^{ref} = z_i^{ref,l} - z_i^{ref,l+1}$ are chosen too large, then E_i will underestimate $E(\nu_i, z_i^{ref})$ too much, whilst if they are too small, the problem will grow in size to the point of becoming computationally intractable.

The objective minimised is given by

$$\sum_{i=1}^{n_z} E_i$$

which attempts to evaluate the economics due to quality by minimising the cost of reprocessing all the off-specification product.

This technique seems promising although it would require a detailed understanding of the probability distribution of the output and of the nature of the quality cost function to give a meaningful and realistic result.

O. Frank also presents a technique for reducing the solution time of the LP for step disturbances. He lets the discrete step response of the ith output of Φ to a positive unit step in the jth disturbance be represented as

$$\sum_{k=1}^{\infty} z^{-k} s_{ij}(k)$$

where the kth element of this discrete step response s_{ij} is given by

$$\begin{aligned} s_{ij}(k) &= \sum_{t=0}^{k} \phi_{ij}(t) \\ &= \left(\sum_{t=0}^{k-1} \phi_{ij}(t)\right) + \phi_{ij}(k) \\ &= \begin{cases} s_{ij}(k-1) + \phi_{ij}(k), & k = 1, ..., \infty \\ \phi_{ij}(0), & k = 0 \end{cases} \end{aligned}$$

giving

$$\phi_{ij}(k) = \begin{cases} s_{ij}(k) - s_{ij}(k-1), & k = 1, ..., \infty \\ s_{ij}(0), & k = 0. \end{cases}$$

Thus any constraint written as

$$b_l \leq \tilde{A}\tilde{x} + \sum_{i=1}^{n_z} \sum_{j=1}^{n_w} \sum_{k=0}^{N_{ij}} A_{ij}^k \phi_{ij}(k) \leq b_u$$

can be rewritten, for a step disturbance in $j = j_{st}$, as

$$b_l \leq \tilde{A}\tilde{x} + \sum_{i=1}^{n_z} \sum_{\substack{j=1 \\ j \neq j_{st}}}^{n_w} \sum_{k=0}^{N_{ij}} A_{ij}^k \phi_{ij}(k) + \sum_{i=1}^{n_z} \sum_{k=0}^{N_{ij_{st}}} A_{ij_{st}}^k (s_{ij_{st}}(k) - s_{ij_{st}}(k-1)) \leq b_u$$

Therefore

$$b_l \leq \tilde{A}\tilde{x} + \sum_{i=1}^{n_z} \sum_{\substack{j=1 \\ j \neq j_{st}}}^{n_w} \sum_{k=0}^{N_{ij}} A_{ij}^k \phi_{ij}(k) + \sum_{i=1}^{n_z} \sum_{k=0}^{N_{ij_{st}}} \bar{A}_{ij_{st}}^k s_{ij_{st}}(k) \leq b_u$$

where

$$\bar{A}_{ij_{st}}^k = \begin{cases} A_{ij_{st}}^k - A_{ij_{st}}^{k-1}, & k = 1, ..., N_{ij} \\ A_{ij_{st}}^0 & k = 0. \end{cases}$$

This allows the parts of the LP associated with the step disturbance $w_{j_{st}}$, i.e., with all $\phi_{ij_{st}}$, to be rewritten in terms of $s_{ij_{st}}$. This is useful, because it means that the condition 3.24 given in section 3.4.2 for the maximum deviation

$$-\nu_{ij_{st}} \leq -\sum_{k=0}^{N_{ij_{st}}} \phi_{ij_{st}}(k) + 2\left(\sum_{k=0}^{t} \phi_{ij_{st}}(k)\right) \leq \nu_{ij_{st}} \quad \text{for} \quad \begin{cases} t = 0, ..., N_{ij_{st}} \\ i = 1, ..., n_z \end{cases}$$

can be restated as

$$-\nu_{ij_{st}} \leq -s(N_{ij_{st}}) + 2s(t) \leq \nu_{ij_{st}} \quad \text{for} \quad \begin{cases} t = 0, ..., N_{ij_{st}} \\ i = 1, ..., n_z \end{cases}$$

This does not reduce the number of constraints, but it does greatly reduce the number of nonzero entries in the constraint matrix ,thus speeding up the solution time of the LP considerably.

3.8 Implementation

The software was implemented as a series of modules programmed in MATLAB and FORTRAN. MATLAB was used to profit from the matrix and control orientated tools and FORTRAN was used to build arrays quickly. The LP was solved using CPLEX .

The problem requires a discrete state-space system described by

$$\begin{aligned} x_{k+1} &= A \quad x_k + [B_1 \ \ B_2] \begin{bmatrix} w_k \\ u_k \end{bmatrix} \\ \begin{bmatrix} z_k \\ y_k \end{bmatrix} &= \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} x_k + \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} w_k \\ u_k \end{bmatrix} \end{aligned} \tag{3.46}$$

plus details on the performance related objective function J and constraints c and a disturbance description for each w_j in the disturbance vector w, i.e.,

$$w_j \in \begin{cases} W_{pr}, \text{ a persistent disturbance } \|w\|_\infty \leq 1 \\ W_{st}, \text{ a step disturbance} \\ W_{ss}, \text{ a steady state disturbance} \end{cases} \tag{3.47}$$

Layouts of the software, set up for both the interpolation conditions and Q-parametrisation techniques, is shown in Figures 3.3 and 3.4 respectively. More details are provided in the appendices indicated.

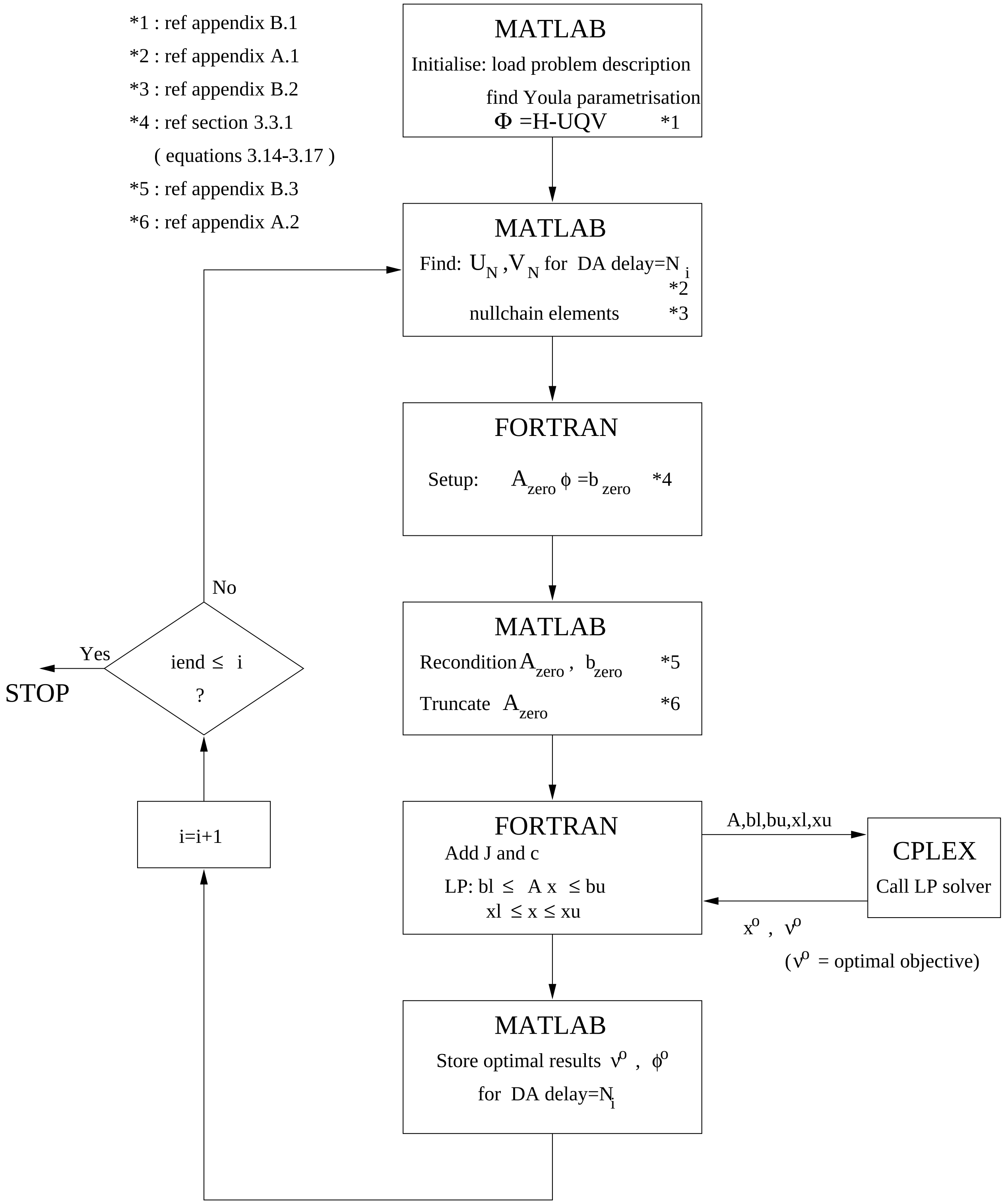

Figure 3.3: Software for the interpolation conditions technique

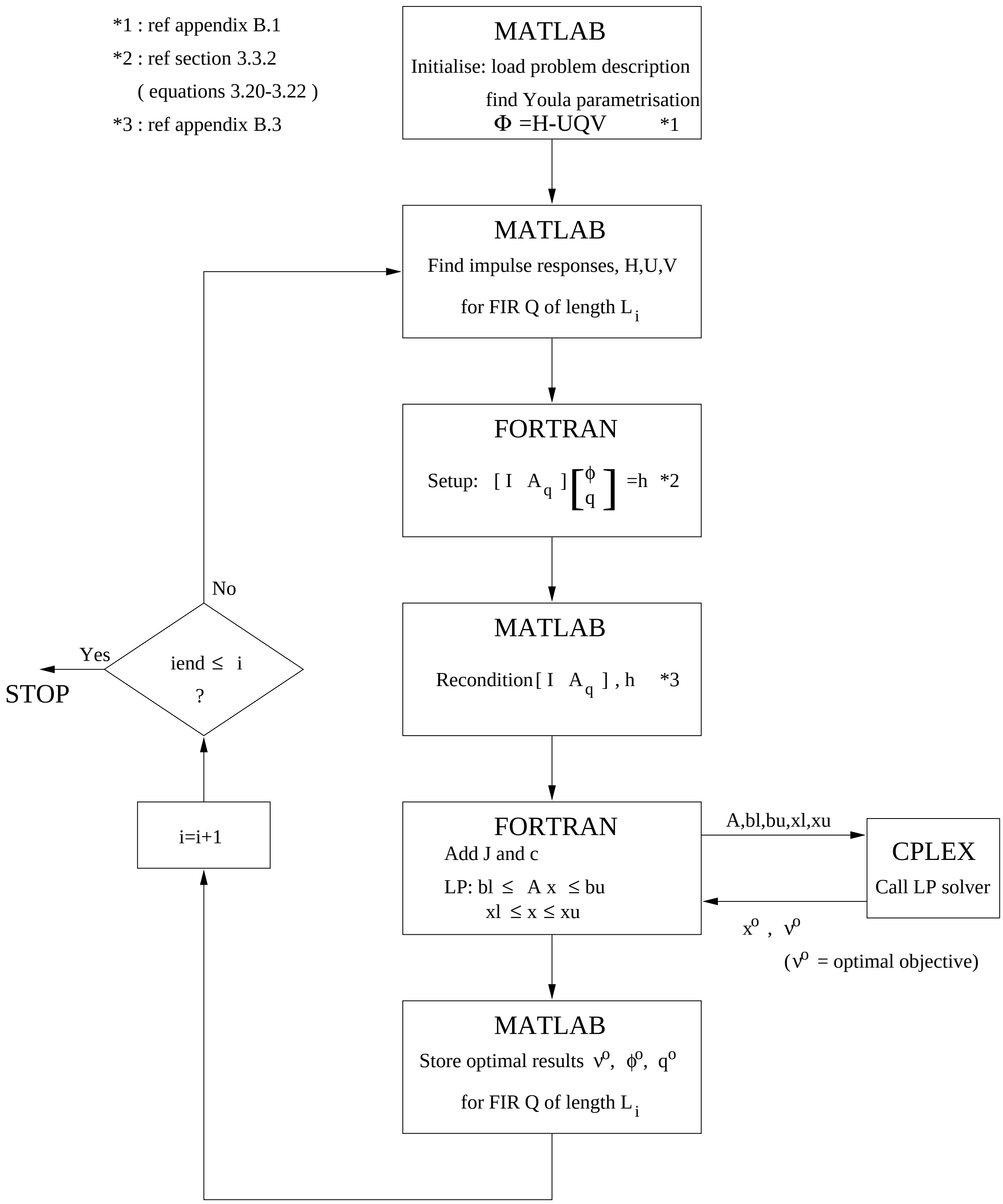

Figure 3.4: Software for the Q-parametrisation technique

The success of these techniques are highly sensitive to the details of the implementation. Therefore how the specific details of the techniques are implemented will directly effect the final computational cost of these methods. The complexity of the implementation itself adds to this cost.

3.9 Review

A linear controllability analysis technique has been developed in the chapter, which is based on the optimal realisable control problem

$$\min_{K,u_0} J(K, u_0) \;\; s.t. \;\; c(K, u_0, w) \leq 0 \;\forall w \in W \tag{3.48}$$

This technique provides an upper (pessimistic) bound on controllability and has the following properties:

- It allows constraints and disturbances to be defined in the time domain. A wide range of typical process performance requirements can be captured, with the restriction that they be expressed as linear functions. Also a range of typical and realistic process disturbance descriptions can be used, i.e., persistent, step or steady state disturbances.

- It optimises the achievable performance over the set of linear time invariant (LTI) controllers, which is a broad and realisable set. There is no added complexity due to using a non-square controller rather than a square controller, thereby allowing a more realistic selection of measured and manipulated variables. Also there is no added complexity due to using both feedforward and feedback control.

- It incorporates most of the fundamental limitations on controllability. The limitation due to NMP characteristics is captured directly by the feasibility constraints. The input constraints can be incorporated by including them in the regulated outputs and enforcing performance constraints on them directly. Any limitation due to measurement noise can be captured by including the noise in the disturbance vector and altering the state space system so that these (noise) disturbances are added directly on to the appropriate measurements. The only fundamental limitation completely ignored in this technique is uncertainty.

- It can be formulated and solved as an LP, therefore it can be used to tackle realistic problems with less computational expense than a convex or nonlinear programming technique.

There are some shortcomings such as the objective and constraints must be linear, the controller set is incomplete, uncertainty is ignored, some problems will give rise to large unwieldy LPs, etc. However, on the whole, this method answers many of the requirements, set out in 2.4.1, for a useful new linear controllability analysis technique. This estimate of controllability is easily interpreted and is most probably more realistic than the optimistic and pessimistic bounds produced by the existing optimal control techniques discussed in 2.3.2. The last point is due to the broad set of realisable controllers used and the fact that the majority of the fundamental limitations on achievable control can be included in the problem.

Conclusions on both this technique and the nonlinear controllability analysis technique developed in the next chapter are drawn in the final chapter of this thesis.

Chapter 4

Nonlinear Controllability Using Optimisation

In the following chapter we present a nonlinear controllability analysis technique, which is complementary to the linear technique presented in Chapter 3. This method assesses the best achievable performance of a nonlinear model for an idealised controller by solving an optimal idealised control problem. The result provides an optimistic bound on the actual best achievable performance. Many fundamental limitations on controller performance are captured by this problem, although not as many as for the linear technique presented in the previous chapter. The performance requirements are no longer limited to linear functions and time domain constraints can be directly incorporated. As for the linear controllability analysis the results are easily interpreted.

This complementary nonlinear technique addresses a similar problem to (3.1)

$$\min_{K,u_0} J(K,u_0) \;\; s.t. \;\; c(K,u_0,w) \leq 0 \; \forall w \in W$$

but for an idealised controller and a nonlinear model. As discussed in 2.4.2 such an optimal idealised control problem gives a lower bound on the controllability. This optimistic bound is given conceptually by

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w(t)\in W} \min_{u(t)\in U} c(u,u_0,w) \leq 0 \tag{4.1}$$

where W is the set of time varying disturbances, within the accepted bounds, and the idealised control set, U, allows any time varying trajectory, $u(t)$, which satisfies the control input constraints. The control schedule $u(t)$ is not included in the objective minimisation since this might suggest that there is one unique control schedule $u(t)$ selected to cope for the complete disturbance set W. In fact a different $u(t)$ is selected for each $w(t) \in W$. On the whole this problem is used to solve for a variable operating point. But for the case when the operating point is constant the constraints become the objective. This is discussed in more detail in section 4.6.

The implemented problem which is solved is given by

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0 \tag{4.2}$$

where $w^p \in \bar{W}$ is given by a specified set of finite step disturbances

$$\bar{W} = \{w^1, w^2, ..., w^p, ..., w^{n_{dist}}\} \tag{4.3}$$

and the control schedule u^p is selected to satisfy the constraints

$$c(u^p, u_0, w^p) \leq 0$$

for the pth disturbance, w^p, in $\bar{W}$. How we arrive at this problem is explained in section 4.1. In contrast to the linear case, the constraints, c, and objective, J, need not be linear functions.

The nonlinear controllability analysis method set up in this chapter is a nonlinear optimal control problem. Therefore this requires both a nonlinear dynamic process model and the use of nonlinear dynamic optimisation techniques to solve the optimisation problem posed in (4.2).

The control is discretised to allow the problem to be solved using nonlinear programming (NLP) techniques. The nature of the idealised controller, including the fact that it is acausal, is discussed. The acausal behaviour of the controller contributes to how optimistic this lower bound on achievable performance is. A technique is presented to try to tighten this lower bound by limiting this acausal element of the control schedules, u^p, selected by the optimiser.

The disturbances $w^p \in \bar{W}$ are described as discretised steps and the subproblem set up for each of these disturbances is presented.

The objectives and constraints, described in the previous chapter, can be transferred to this nonlinear problem with no difficulty. Both linear and nonlinear objectives and constraints can be used.

Finally, how the problem is formulated and solved as a NLP is presented. The nonlinear dynamic optimisation techniques required to solve the NLP are typically computationally very expensive

The specialisation of the nonlinear controllability technique to linear models has been developed. This linear version can be solved efficiently using linear programming and avoids the computational expense of the nonlinear method. Some details, specific to this linear method, are discussed and a simple linear example is used to demonstrate the technique for limiting the acausal behaviour of the idealised controller.

4.1 The Implemented Problem

The conceptual problem that we wish to solve is given by

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w(t) \in W} \min_{u(t) \in U} c(u, u_0, w) \leq 0$$

where W and U are the sets of time varying disturbances and control inputs, respectively, within their accepted bounds.

For this problem to be implementable, both u and w must be parametrised, as discussed in 4.3 and 4.4. The parametrisation of w is simple when restricted to steps, as in the technique presented in this chapter, otherwise it introduces a potential optimism to the measure, since the parametrised set will be restricted compared to the original signal. However, since this result is already a lower bound, this added optimism would be acceptable, simply making it a lower bound on a lower bound. For this technique, u is parametrised, such that it is discrete and piecewise linear, and therefore is limited only by the sampling rate chosen. There are good heuristics for the selection of this rate and it can be adjusted to check the sensitivity of the result to the sampling.

This limitation on the description of u is both realistic and consistent with the linear technique presented in the last chapter.

However this problem is still difficult to solve, therefore to simplify it the set of disturbances, W, is restricted to a specified finite set, $\bar{W}$, of steps

$$\bar{W} = \{w^1, w^2, ..., w^p, ..., w^{n_{dist}}\}$$

This adds further optimism to the result, meaning that it is still a lower bound on the controllability. Thus the problem becomes

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

where w^p and u^p are discretised over time. Therefore a set of control schedules, $u^p, p = 1, .., n_{dist}$, are selected such that u^p satisfies the constraints

$$c(u^p, u_0, w^p) \leq 0$$

for the pth disturbance, w^p, in $\bar{W}$. Let this finite set of control schedules, $u^p, p = 1, .., n_{dist}$, be described by $\bar{U} = \{u^1, u^2, ..., u^p, ..., u^{n_{dist}}\}$.

4.2 The Nonlinear Dynamic Process Model

The nonlinear controllability analysis method, set up in this chapter, is a nonlinear optimal control problem described by (4.2). This requires a nonlinear dynamic process model, relating u and w to the output variables $y(t)$ and $x(t)$ involved in the performance constraints and the performance objective.

In general dynamic process models are described by sets of differential algebraic equations (DAE's)

$$F(\dot{x}, x, y, \theta, t) = 0 \tag{4.4}$$

with initial conditions $x(t_0)$, where $x(t)$ are the state variables, $y(t)$ are the algebraic variables and θ are parameters such as control variables, $u(t)$, or time invariant parameters, v. In the case of nonlinear dynamic process models F is a nonlinear function.

For the optimal control problem, tackled in this chapter, we require both the control function $u(t)$ and the disturbance $w(t)$ to be parametrised as explicit functions of time, using a finite number of parameters, as discussed in 4.3 and 4.4. This controllability analysis technique is based on the optimisation of the performance in the face of a set of disturbances, $w^p \in \bar{W}$, by the selection of a set of control functions, $u^p \in \bar{U}$. Therefore we must include both the control input u and the disturbance w in the description (4.4)

$$F(\dot{x}, x, y, u, w, t) = 0 \tag{4.5}$$

This controllability analysis always starts from steady state, giving the initial conditions

$$\dot{x}(t_0) = 0. \tag{4.6}$$

Therefore for a given disturbance, $w^p(t) \in \bar{W}$, and a corresponding optimal control input, $u^p(t) \in \bar{U}$, we have the equality constraints in (4.7) which give the variables $y^p(t)$ and $x^p(t)$.

$$\left.\begin{array}{l} F(\dot{x}^p, x^p, y^p, u^p, w^p, t) = 0 \text{ for } u^p \in \bar{U},\ w^p \in \bar{W} \\ \dot{x}^p(t_0) = 0 \end{array}\right\} \tag{4.7}$$

4.3 Control Parameterisation

As it stands, the nonlinear optimal control problem given in (4.2), for the model described in (4.7), could not be tackled using nonlinear programming, since, for every value of t, the control variable $u^p(t)$ can take a different value. Since there are an infinite number of values of t within this time interval this gives an infinite dimensional optimisation problem. To solve this using nonlinear programming, we must transform the problem to a finite dimensional problem.

To do this we can parameterise the control function as described in (Edgar and Himmelblau, 1988)

$$u^p(t) = \sum_{i=0}^{p} a_i \Theta_i(t) \tag{4.8}$$

where Θ_i is a specified function of time, such as t^i, which would make u^p a pth order polynomial,

$$u^p(t) = a_0 + a_1 t + a_2 t^2 + ... + a_p t^p \tag{4.9}$$

and it is the finite number of coefficients, a_i, that are selected as the optimisation variables. The time horizon $[t_0, t_f]$ can be broken into n time intervals and different a_i's selected for each interval, i.e., the a_i's are discretised into $a_i(0)$, $a_i(1)$,.... $a_i(n-1)$. If $p = 0$ and $\Theta_0 = 1$, then this is simply the discretisation of the control function into piecewise constant values of u^p, i.e., $u^p(0)$, $u^p(1)$,.... $u^p(n-1)$. This gives a finite dimensional problem, which can be solve using NLP techniques.

For the nonlinear controllability analysis method developed here, we would discretise the control function with the same sampling period as used for the linear controllability analysis technique, discussed in Chapter 3, so to maintain consistency. This gives

$$u^p = [u^p(0), u^p(1), ...u^p(n-1)]. \tag{4.10}$$

where the control trajectory is piecewise linear between the discrete values $u^p(k)$ and $u^p(k+1)$.

Each control schedule u^p must start from the same steady state

$$u_0 = u^1(0) = u^2(0) = ... = u^{n_{dist}}(0) \tag{4.11}$$

to ensure that the performance over all the disturbances in $\bar{W}$ is assessed for a common operating point u_0.

4.4 The Disturbance Description

This controllability analysis assesses the best achievable performance for a set of step disturbances, $\bar{W}$, as described in (4.2)

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

This set $\bar{W}$ is user specified. If the process is well understood, then the typical process disturbances may have been identified, in which case these can be put into the set $\bar{W}$.

Otherwise, the linear controllability analysis in Chapter 3 can be used to give estimates of the worst step disturbances for this problem.

The work in this chapter concentrates on step disturbances, since these are common in process systems and can be described exactly by a finite number of parameters. However this analysis could be extended to other types of disturbances.

As mentioned in section 4.2, the disturbance, w^p, must be parametrised as an explicit function of time using a finite number of parameters. Unlike the control input, u^p, which is selected by the optimisation, the disturbance, w^p, is user supplied. Since the disturbances are being chosen from the set of step disturbances, it is simple to describe them exactly using such a parametrisation.

A step disturbance, w^p, can be described as

$$w^p = \begin{bmatrix} w_1^p \\ w_2^p \\ \vdots \\ w_{n_w}^p \end{bmatrix}, \quad w_j^p(t) = \left\{ \begin{array}{cc} \text{its steady state value,} & \text{if } t < t_j^p \\ \text{a new value } sign_j^p & \text{if } t \geq t_j^p \end{array} \right\} j = 1, ..., n_w \quad (4.12)$$

where $sign_j^p$ is the value that w_j^p steps to at time t_j^p. So w^p can be discretised as piecewise constant

$$w^p = [w^p(0), w^p(1), ...w^p(n-1)] \quad (4.13)$$

with a corresponding set of discrete times, $t^p(k)$ $k = 1, ..., n-1$, chosen so that the disturbance has constant value $w^p(k)$ for the time period $t^p(k) \leq t < t^p(k+1)$. If the jth component w_j^p of the disturbance w^p steps to $sign_j^p$ at time t_j^p then the discrete elements describing w_j^p

$$w_j^p = [w_j^p(0), w_j^p(1), ...w_j^p(n-1)]$$

are given by

$$w_j^p(k) = \left\{ \begin{array}{cc} \text{its steady state value,} & \text{if } t^p(k) < t_j^p \\ sign_j^p & \text{if } t(k) \geq t_j^p \end{array} \right.$$

where $t(k)$ should be chosen so that for every j, $\exists k_j$, such that $t(k_j) = t_j^p$ and $\min_l t_l^p$ should be set to 0.

All the step disturbances in $\bar{W}$ should start from the same steady state w_0, i.e.,

$$w_0 = w^1(0) = w^2(0) = ... = w^{n_{dist}}(0) \tag{4.14}$$

so that, together with u_0, this defines a common operating point at which the performance is evaluated. On the whole, this disturbance level w_0 should be chosen to be the expected disturbance level $\bar{w}$, as for the ONDE (Optimal Nonlinear Dynamic Economic) problem. Note that, unlike the OLDE technique, choosing $w_0 = \bar{w}$ does not mean that the expected value of the objective is given, since neither the model nor the idealised controller are linear (see 4.6.1).

Putting (4.7), (4.11) and (4.16) together with the constraint, $c(u^p, u_0, w^p) \leq 0$, from (4.2), we get a feasibility subproblem for each disturbance $w^p \in \bar{W}$.

$$\left.\begin{array}{c} c(x^p, y^p, u^p, u_0) \leq 0 \\ F(\dot{x}^p, x^p, y^p, u^p, w^p, t) = 0 \\ \dot{x^p}(t_0) = 0 \\ u^p(0) = u_0 \\ u^l \leq u^p \leq u^h \end{array}\right\} \tag{4.15}$$

If this subproblem is not satisfied for u^p, then this means that u^p is not a feasible control input for disturbance w^p.

4.5 The Idealised Controller

The optimal controller, selected for the nonlinear controllability problem, is the control schedule given in (4.10). This control schedule is selected to optimise the objective and to satisfy the constraints for each disturbance $w^p \in \bar{W}$, including constraints on itself.

$$u^l \leq u^p \leq u^h \quad \text{for } p = 1, ...n_{dist} \tag{4.16}$$

Since it is not an implementable feedback or feedforward controller and might not be achievable in reality, it describes an idealised controller. Specifically, this is due to three reasons. The first being that limitations, due to the measured variables available, are not imposed on the control performance. The second being that the control schedule is selected to optimise the performance, as captured by the objective, over the finite horizon $[t_0, t_f]$, without having to assure any kind of performance beyond this horizon. The last being that the optimisation has full information of the disturbance, even before it occurs, which means that the controller is acausal ,i.e, uses knowledge of the future behaviour of the disturbance.

The idealised nature of the optimal control, means that the result, provided by the solution of problem 4.2, provides an optimistic bound on the best achievable performance of the process. It is unlikely that this bound will be achievable in reality, since the set of controllers supplied to the optimal control problem is much larger than the set of implementable controllers available in reality. It is important to make this bound as realistic as possible by limiting these idealised elements of the control as much as possible, so that the set of controllers, for the optimisation, more closely represents the set of implementable controllers.

Therefore a technique for limiting the acausal nature of the optimal controller for the nonlinear controllability analysis is proposed in section 4.5.1. Details specific to the implementation of this for the linear specialisation of the nonlinear controllability technique are presented in section 4.8.2.

The optimistic nature of the result of this controllability analysis, means that if the process fails to be feasible for this test, then it suggests it will be unable to achieve it with any real, implementable controller. The conclusion of such a test, can only be used as a suggestion, rather than a proof of infeasibility, due to implementation issues, such as the accuracy of the nonlinear model, the discretisation of the controller and whether the global optimum has been found.

4.5.1 Limiting Acausal Behaviour

One way to limit the acausal element of the controller is to ensure that the control schedule does not move until a disturbance has appeared. However, if the disturbance w, for which the performance must be evaluated, includes more than one component

$w = [w_1, w_2, ...w_{n_w}]^T$ and all these components do not step simultaneously, then although we can hold u still until the first occurrence of a step, it must then be freed to respond before the steps in all the other components have appeared. This means, that the optimiser can select the optimal control schedule to best prepare for the future appearance of steps, which it should not know about, making it acausal.

Another aspect of this optimistic bound is that time delays, which exist between the occurrence of a disturbance and the measurement of it, are not imposed on the idealised controller. If feedforward control can be used for the problem, then this aspect is not an issue.

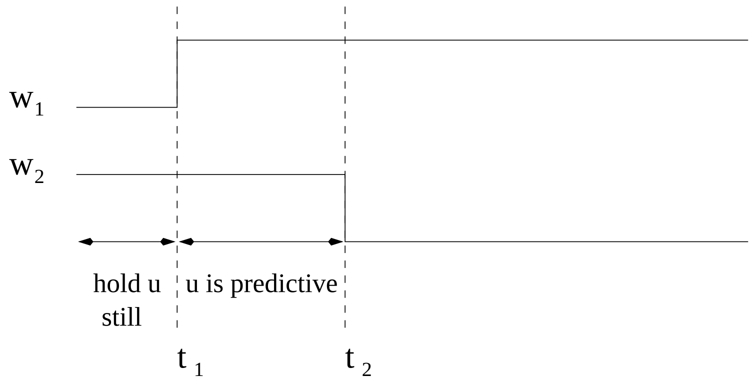

Figure 4.1: An example disturbance that allows acausal control

Therefore, a technique for limiting the controllers ability to take advantage of its knowledge of future disturbances, when selecting a control schedule, would be useful in providing a tighter optimistic bound on the achievable performance.

For the problem in (4.2)

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

the constraints must be satisfied over the user specified set of n_{dist} step disturbances

$\bar{W}$. Each disturbance, $w^p \in \bar{W}$, has n_w components, i.e.,

$$w^p = \begin{bmatrix} w_1^p \\ w_2^p \\ \vdots \\ w_{n_w}^p \end{bmatrix}. \tag{4.17}$$

For a particular disturbance, $w^{\breve{p}} \in \bar{W}$, the jth component can be described as

$$w_j^{\breve{p}}(k) = \begin{cases} \text{its steady state value,} & \text{if } t^{\breve{p}}(k) < t_j^{\breve{p}} \\ sign_j^{\breve{p}} & \text{if } t(k) \geq t_j^{\breve{p}} \end{cases} \tag{4.18}$$

where $sign_j^{\breve{p}}$ is the value that $w_j^{\breve{p}}$ steps to at time $t_j^{\breve{p}}$. In the following an assumption is made that the performance is worst for either the greatest positive or greatest negative step change, so $sign_j^{\breve{p}}$ will either be the upper bound, w_j^h, or the lower bound, w_j^l, on the jth component of the disturbance.

If the following technique is used to limit the acausal behaviour of the controller, then a further n_{dist} sets of step disturbances, given by $\bar{W}^p, p = 1..., n_{dist}$, are added to the problem along with the set, $\bar{W}$, of n_{dist} user specified step disturbances. Each set $\bar{W}^p$ correlates to the disturbance $w^p \in \bar{W}$. Therefore, for a particular disturbance $w^{\breve{p}} \in \bar{W}$, the set $\bar{W}^{\breve{p}}$ would be selected to limit the acausal behaviour of the idealised controller for this specific disturbance.

If all the steps $w_j^{\breve{p}}$ occur at the same time, they should be described so that $t_j^{\breve{p}} = 0 \; \forall j$, in which case the set $\bar{W}^{\breve{p}}$ will be empty, i.e., $\bar{W}^{\breve{p}} = \emptyset$. Otherwise, for each component $w_j^{\breve{p}}$, we check if

$$t_j^{\breve{p}} > \min_l t_l^{\breve{p}}, \tag{4.19}$$

in which case, the controller will have been acausal over the interval

$$[\min_l t_l^{\breve{p}}, t_j^{\breve{p}}].$$

with regards to the disturbance component $w_j^{\breve{p}}$. This means, that using the following technique, we will search for a disturbance $\tilde{w}^{\breve{p}j}$, which will limit this acausal behaviour. This is done for each jth component, $w_j^{\breve{p}}$, individually and the set $\bar{W}^{\breve{p}}$ is made up of all the successfully found $\tilde{w}^{\breve{p}j}$'s for $j = 1, ...n_{dist}$.

The procedure for finding $\tilde{w}^{\breve{p}j}$, for each component $w_j^{\breve{p}}$, is described in the following and is based on the principle that

$$\text{if } \tilde{w}^{\breve{p}j}(t) = w^{\breve{p}}(t) \text{ for } t \leq t_j^{\breve{p}}$$
$$\text{then causality implies that } \tilde{u}^{\breve{p}j}(t) = u^{\breve{p}}(t) \text{ for } t \leq t_j^{\breve{p}}$$

where $\tilde{u}^{\breve{p}j}$ and $u^{\breve{p}}$ are the control schedules corresponding to the disturbances $\tilde{w}^{\breve{p}j}$ and $w^{\breve{p}}$ respectively.

Let the optimal solution to the original problem

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

be described by J^o, u_0^o and $u^{p,o}, p = 1, ..., n_{dist}$.

Splitting the control schedules into u_b and u_a, so that

$$u(t) = \begin{cases} u_b(t), & t < t_j^{\breve{p}} \\ u_a(t), & t \geq t_j^{\breve{p}} \end{cases} \tag{4.20}$$

it is $u_b^{\breve{p},o}$ which has made use of the its future knowledge of $w_j^{\breve{p}}$ to improve J^o. The idea proposed to limit the acausal behaviour of $u_b^{\breve{p}}$ with respect to the disturbance component $w_j^{\breve{p}}$, is to find a disturbance, $\tilde{w}^{\breve{p}j}$, which differs from $w^{\breve{p}}$ in its behaviour beyond $t_j^{\breve{p}}$,

$$\tilde{w}^{\breve{p}j}(t) = w^{\breve{p}}(t) \quad t < t_j^{\breve{p}} \tag{4.21}$$

such that it causes a violation of the constraints with any control schedule beginning with $u_b^{\breve{p},o}$. This forces the controller $\tilde{u}^{\breve{p}j}$ to select a different control schedule for $t < t_j^{\breve{p}}$ to fulfill the constraints, i.e.,

$$c(\tilde{u}^{\breve{p}j}, \tilde{u}_0, \tilde{w}^{\breve{p}j}) \leq 0 \Rightarrow \tilde{u}_b^{\breve{p}j} \neq u_b^{\breve{p},o} \tag{4.22}$$

It may well be that $\tilde{u}_0 \neq u_0^o$.

The choice of this disturbance is limited by the constraint in (4.21). This means that the two disturbances must not differ until the step in $w_j^{\breve{p}}$ occurs. However, after this point, the behaviour of $\tilde{w}^{\breve{p}j}$ should be selected so that its optimal acausal control behaviour, for the period $t < t_j^{\breve{p}}$, would be different to that chosen for $w^{\breve{p}}$ for this same period. The controller is then presented with these two possible future disturbance scenarios for which it must select an initial control schedule that will cope equally well for either. This, therefore, reduces its ability to use its future knowledge of $w^{\breve{p}}$ to select an optimal acausal control schedule for this scenario alone.

The possible choices of the hth component of the disturbance $\tilde{w}^{\breve{p}j}$ are given by

$$\tilde{w}_h^{\breve{p}j}(t) = \begin{cases} w_h^{\breve{p}}(t) & t < t_h^{\breve{p}j} \\ \left.\begin{array}{ll} w_h^{\breve{p}}(t) & \text{if } t_h^{\breve{p}} < t_j^{\breve{p}} \\ sign_h^{\breve{p}j} & \text{if } t_h^{\breve{p}} \geq t_j^{\breve{p}} \end{array}\right\} & t \geq t_h^{\breve{p}j} \end{cases} \tag{4.23}$$

where $sign_h^{\breve{p}j}$ can be chosen to be w_h^h or w_h^l and $t_h^{\breve{p}j} \geq t_j^{\breve{p}}$, thus defining the step. Note that if the step in $w_h^{\breve{p}}$ occurs before $t_j^{\breve{p}}$, then the hth component of $\tilde{w}^{\breve{p}j}$, $\tilde{w}_h^{\breve{p}j}$, will be left the same as $w_h^{\breve{p}}$ for all time.

Let us split the disturbances in the same manner as the controls, i.e.,

$$w(t) = \begin{cases} w_b(t), & t < t_j^{\breve{p}} \\ w_a(t), & t \geq t_j^{\breve{p}} \end{cases} \tag{4.24}$$

The set of possible choices for $\tilde{w}_a^{\breve{p}j}$ is described by $\tilde{W}_a$, where $\tilde{w}_a^{\breve{p}j}$ is given by 4.23 for $t \geq t_j^{\breve{p}}$ and all the possible choices of $sign_h^{\breve{p}j} = w_h^h$ or w_h^l and $t_h^{\breve{p}j} \geq t_j^{\breve{p}}$.

Find the disturbance which forces the maximum change in $J(u_0)$. To force such a change there must be a $\tilde{w}_a \in \tilde{W}_a$ such that $\nexists\ \tilde{u}_a \in U$ which satisfies

$$\max_k c_k(\tilde{u}, \tilde{u}_0, \tilde{w}) \leq 0, \quad \tilde{u}_b = u_b^{\breve{p},o} \text{ and } \tilde{w}_b = w_b^{\breve{p}}.$$

To find such a disturbance,$\tilde{w}^{\breve{p}j}$, the following optimisation could be solved which would force a change in $J(u_o)$.

$$\tilde{J}(\tilde{w}_a, \tilde{u}_a) = \max_{\tilde{w}_a \in \tilde{W}_a} \min_{\tilde{u}_a \in U} \max_k c_k(\tilde{u}, u_0, \tilde{w}) \;\; s.t. \;\; \tilde{u}_b = u_b^{\breve{p},o}, \;\; \tilde{w}_b = w_b^{\breve{p}} \tag{4.25}$$

where c_k is the kth constraint in c. If the optimal $\tilde{J}^o > 0$, then this means that with optimal disturbance $\tilde{w}^o$ the constraints $c(\tilde{u}, \tilde{u}_0, \tilde{w}^o) \leq 0$ cannot be satisfied for any control schedule $\tilde{u}$ with $\tilde{u}_b = u_b^{\breve{p},o}$. Therefore since $\tilde{w}^o$ satisfies the condition (4.22), we let $\tilde{w}^{\breve{p}j} = \tilde{w}^o$.

However this is a max-min problem and cannot be solved efficiently for general cases. To convert this problem to a conventional maximisation problem, we simplify it by defining a finite set of possible control schedules, $\tilde{U}^{\check{p}j}$, over which the constraints are minimised.

$$\tilde{u}^p \in \tilde{U}^{\check{p}j} \Rightarrow \tilde{u}^p(t) = \begin{cases} u_b^{\check{p},o}(t), & t < t_j^{\check{p}} \\ u_a^{p,o}(t), & t \geq t_j^{\check{p}} \end{cases} \quad p = 1, \ldots n_{dist} \tag{4.26}$$

The problem solved instead of (4.25) is

$$\tilde{J}_1(\tilde{w}_a) = \max_{\tilde{w}_a \in \tilde{W}_a} \min_{\tilde{u}^p \in \tilde{U}^{\check{p}j}} \max_k c_k(\tilde{u}^p, u_0, \tilde{w}) \;\; s.t. \;\; \tilde{w}_b = w_b^{\check{p}} \tag{4.27}$$

If $\tilde{J}_1^o > 0$, then a disturbance $\tilde{w}^o$ has been found that makes the constraints $c_k(\tilde{u}^p, u_0, \tilde{w}^o)$ infeasible $\forall \tilde{u}^p \in \tilde{U}^{\check{p}j}$. Therefore $\tilde{w}^o$ is a candidate disturbance for $\tilde{w}^{\check{p}j}$. To check if it satisfies the condition given in (4.22) solve

$$\tilde{J}_2(\tilde{u}_a) = \min_{\tilde{u}_a \in U} \max_k c_k(\tilde{u}, u_0, \tilde{w}^o) \;\; s.t. \;\; \tilde{u}_b = u_b^{\check{p},o}. \tag{4.28}$$

If the optimal $\tilde{J}_2^o > 0$, then $\tilde{w}^o$ satisfies (4.22), i.e.,

$$c(\tilde{u}^{\check{p}j}, \tilde{u}_0, \tilde{w}^o) \leq 0 \Rightarrow \tilde{u}_b^{\check{p}j} \neq u_b^{\check{p},o}$$

and $\tilde{w}^{\check{p}j} = \tilde{w}^o$.

In a similar way, for every step disturbance $w^p \in \bar{W}$ check the step time, t_j^p, of each w_j^p and if

$$t_j^p > \min_l t_l^p,$$

then find the corresponding $\tilde{w}^{pj}$. Let these $\tilde{w}^{pj}$ for $w^p \in \bar{W}$ make a finite set $\bar{W}^p$. If for some w^p all the components step simultaneously $t_j^p = 0 \;\; \forall j$, then $\bar{W}^p = \emptyset$.

To find the optimal objective, J, with the acausal element of the controller limited, solve the original problem with some added constraints.

$$\begin{aligned} \min_{u_0} J(u_0) \;\; s.t. \quad & \forall w^p \in \bar{W} \\ & \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0 \\ & \max_{\tilde{w}^{pj} \in \bar{W}^p} \min_{\tilde{u}^{pj} \in U} c(\tilde{u}^{pj}, u_0, \tilde{w}^{pj}) \leq 0 \\ & \tilde{u}_b^{pj} = u_b^p \end{aligned} \tag{4.29}$$

We may not always be able to find a suitable w^{pj} with this technique. This is highly problem specific. Also this technique cannot guarantee causality, simply that the optimistic bound it produces will be tighter than that produced by the original acausal problem.

When the problem has a constant operating point, u_0, e.g., the maximum deviation is being minimised, then there are no constraints, as such, since the constraints have become the objective. In this case, if the optimal objective of the original acausal problem is J^o, then the constraints in (4.22) can be given as

$$J(\tilde{u}^{\check{p}j}, \tilde{w}^{\check{p}j}) - J^o \leq 0 \Rightarrow \tilde{u}_b^{\check{p}j} \neq u_b^{\check{p},o}$$

A simple example of this technique for a linear model is presented in section 4.8.3.

This technique can be slightly altered to try to restrict the idealised controller to causal feedback control, i.e., to limit the controllers ability to use feedforward. A similar approach as that described above could be used, but based on the principle

$$\begin{gathered}\text{if } y(\tilde{w}^{\check{p}j})(t) = y(w^{\check{p}})(t) \text{ for } t \leq t_j^{\check{p}} + t_{d_{w_j,y}} \\ \text{then causal feedback implies that } \tilde{u}^{\check{p}j}(t) = u^{\check{p}}(t) \text{ for } t \leq t_j^{\check{p}} + t_{d_{w_j,y}}\end{gathered}$$

where $\tilde{u}^{\check{p}j}$ and $u^{\check{p}}$ are the control schedules corresponding to the measurements $y(\tilde{w}^{\check{p}j})$ and $y(w^{\check{p}})$ of the disturbances $\tilde{w}^{\check{p}j}$ and $w^{\check{p}}$ respectively. It is assumed that the jth component of the disturbance is not measured until the minimum time delay $t_{d_{w_j,y}}$, which is given by

$$t_{d_{w_j,y}} = \min_i t_{d_{w_j,y_i}}$$

where $t_{d_{w_j,y_i}}$ is the estimated time delay from w_j to the measurement y_i.

In this case, the possible choices of the hth component of the disturbance $\tilde{w}^{\check{p}j}$, would be given by the selections of $sign_h^{\check{p}j} = w_h^h$ or w_h^l and $t_h^{\check{p}j} \geq t_j^{\check{p}} + t_{d_{w_j,y}} - t_{d_{w_h,y}}$ in

$$\tilde{w}_h^{\check{p}j}(t) = \begin{cases} w_h^{\check{p}}(t) & t < t_h^{\check{p}j} \\ \left.\begin{array}{ll} w_h^{\check{p}}(t) & \text{if } t_h^{\check{p}} + t_{d_{w_h,y}} < t_j^{\check{p}} + t_{d_{w_j,y}} \\ sign_h^{\check{p}j} & \text{if } t_h^{\check{p}} + t_{d_{w_h,y}} \geq t_j^{\check{p}} + t_{d_{w_j,y}} \end{array}\right\} & t \geq t_h^{\check{p}j} \end{cases} \tag{4.30}$$

and the initial control profile would be forced to be

$$\tilde{u}^{\breve{p}j}(t) = u^{\breve{p}}(t) \quad \text{for} \;\; t < t_j^{\breve{p}} + t_{d_{w_j},y}. \tag{4.31}$$

Further work would be needed to implement and demonstrate this addition to the acausality limiting technique.

4.6 Performance Requirements

To assess the controllability of a nonlinear process model, using the method described in (4.2),

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

we need to define the objective function, J, and the constraints, c, which capture the specific performance requirements for a problem. These can be linear or nonlinear functions of the problem variables, i.e., x^p, y^p and u^p for $p = 1, ..n_{dist}$ and u_0.

The constraints can be used to enforce performance constraints, such as variable magnitude bounds, directly on the problem variables. If there are n_z constrained variables in x and y let these be denoted by z, in which case the variable bounds for the pth disturbance ($p = 1, n_{dists}$) are

$$z_i^l \leq z_i^p \leq z_i^h \quad \text{for } i = 1, ...n_z, \tag{4.32}$$

Therefore the constraints $c(u^p, u_0, w^p) \leq 0$ for each $p = 1, n_{dists}$ would be given by

$$\left.\begin{array}{rcl} z_i^l - z_i^p & \leq & 0 \\ z_i^p - z_i^h & \leq & 0 \end{array}\right\} \forall i$$

The control input bounds in (4.16) are implemented as bounds on the optimisation variables, since u^p are the optimisation variables.

If we wish to assess the maximum deviation in the constrained variables z, to see whether the constrained variables would violate their bounds for the disturbance set $\bar{W}$, then the operating point, u_0, becomes a constant and is no longer an optimisation variable. In this case, the constraints themselves become the objective and are minimised. The maximum deviation for each disturbance can be evaluated as

$$\nu^p = \max_i \max_t (|z_i^p(t) - z_i(0)|) \tag{4.33}$$

So the objective of the nonlinear optimisation problem to be solved is given by

$$\min_{\bar{U}} \max_p \nu^p \tag{4.34}$$

where the finite set of control schedules $\bar{U} = \{u^1, u^2, ..., u^p, ..., u^{n_{dist}}\}$ is chosen, so that each u^p satisfies its subproblem (4.15). If we have a reasonable basis for penalising the maximum deviation on each z_i, then a weighted sum of the deviations, ν_i, given by

$$\nu_i = \max_p \max_t (|z_i^p(t) - z_i(0)|) \tag{4.35}$$

can be minimised. Clearly the objective can be as flexible as for the linear problem, whilst also allowing the use of nonlinear objectives if so desired.

4.6.1 The Optimal Nonlinear Dynamic Economic (ONDE) Problem

If, as for the linear problem, we wish to assess the economic performance of the system, then we set up a similar problem to the OLDE problem for the nonlinear model. As discussed in 3.5.2, most process design involves a steady state minimisation of some objective function o, which is chosen to provide a measure of economic performance, i.e., the cost of operation, the loss in profit. The optimal economic operating point, given by this, will generally require several of the process variables to operate on their inequality constraints. To avoid violation of these constraint,s due to the appearance of process disturbances, the process must operate at a point backed off from this optimal. The aim of the ONDE technique, as for the OLDE technique, is to estimate the minimum back off required, with optimal idealised control, to accommodate the process disturbances without violating constraints.

A difference to the OLDE technique is that the optimal expected value of the objective, over the disturbance set, cannot be assessed, since for a nonlinear objective function o, $E(o(w)) = o(E(w))$, $E(w) = \bar{w}$ no longer holds. Although it still makes sense to select the disturbance operating point, w_0, about which the step disturbances are applied, to be the same as the expected disturbance level.

The objective for the ONDE analysis is the value of the objective function o at the operating point described by the steady state of the disturbance variables before the steps are applied, w_0, and the operating point for the control variables ,u_0. Therefore problem (4.2) becomes

$$\min_{u_0} o(w_0, u_0) \quad s.t. \quad \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0 \tag{4.36}$$

where u_0 gives the new backed off operating point, which ensures that the variables z don't violate their constraints (4.32), whilst at the same time minimising the economic objective o.

4.7 Solving as a Nonlinear Program (NLP)

Nonlinear optimal control problems can be solved using nonlinear programming (NLP) techniques, which optimise an objective function subject to a set of constraints, where any of these constraints or objective may be nonlinear functions, i.e.,

$$\min_{\mathrm{x}} f(\mathrm{x}) \tag{4.37}$$

$$\text{subject to } h_j(\mathrm{x}) = 0 \quad j = 1, 2, ..., meq \tag{4.38}$$

$$g_j(\mathrm{x}) \geq 0 \quad j = meq + 1, ..., m \tag{4.39}$$

$$\mathrm{x}_l \leq \mathrm{x} \leq \mathrm{x}_u \tag{4.40}$$

where x is the vector of the optimisation variables.

For a nonlinear optimal control the objective function is minimised, subject to constraints, over the time interval $[t_0, t_f]$ and the optimisation variables are the parametrised or discretised control variables u. Such a problem can be set up to tackle a wide range of optimisation problems, depending on specific choices of the optimisation variables, constraints and objective function.

The controllability analysis proposed in this chapter, for nonlinear models, requires the solution of the problem in (4.2). The control variables have been discretised to give (4.10)

$$u^p = [u^p(0), u^p(1), ...u^p(n-1)]$$

Therefore the optimisation variable x consists of $\bar{U} = \{u^1, ...u^{n_{dist}}\}$ (the common operating point u_0 is given by $u^p(0)$).

The feasibility subproblem for disturbance w^p, given in (4.15), gives a set of equality constraints

$$h^p(\mathrm{x}) = 0 \equiv \begin{cases} \dot{\mathrm{x}}^p(t_0) = 0 \\ F(\dot{x}^p, x^p, y^p, u^p, w^p, t) = 0 \end{cases} \tag{4.41}$$

and a set of inequality constraints

$$g^p(\mathrm{x}) \geq 0 \equiv -c(x^p, y^p, u^p) \geq 0 \tag{4.42}$$

where $u^p(0) = u_0$ has been substituted to drop u_0.

These constraints are stacked together for all the disturbances in $\bar{W}$, to give the equality constraints

$$h(x) = \begin{bmatrix} h^1(\mathrm{x}) \\ h^2(\mathrm{x}) \\ \vdots \\ h^{n_{dist}}(\mathrm{x}) \\ h^{u_0}(\mathrm{x}) \end{bmatrix} = 0 \tag{4.43}$$

and the inequality constraints

$$g(x) = \begin{bmatrix} g^1(\mathrm{x}) \\ g^2(\mathrm{x}) \\ \vdots \\ g^{n_{dist}}(\mathrm{x}) \end{bmatrix} \geq 0 \tag{4.44}$$

where $h^{u_0}(\mathrm{x}) = 0$ are an extra $n_u(n_{dist} - 1)$ equality constraints given by

$$h^{u_0}(\mathrm{x}) = 0 \equiv u_j^p(0) - u_j^{p-1}(0) = 0, \quad \text{for} \begin{cases} p = 2, ...n_{dist} \\ j = 1, ...n_u \end{cases} \tag{4.45}$$

which enforce a common operating point as described in (4.11).

Finally the objective function is simply given by

$$f(\mathrm{x}) \equiv J(u^p(0)) \tag{4.46}$$

where $u^p(0)$ has been substituted for u_0 again and the optimisation variables constraints are given by (4.16), i.e.,

$$\mathrm{x}_l \leq \mathrm{x} \leq \mathrm{x}_u \equiv u^l \leq u^p \leq u^h \quad \text{for } p = 1, ...n_{dist} \tag{4.47}$$

There are several ways to solve NLP problems with constraints such as this: lagrange multiplier methods, iterative linearisation methods, penalty function methods and iterative quadratic programming. However we only discuss the last method, since this is what we have used to implement the nonlinear controllability analysis. The presentation of the successive quadratic programming (SQP) method for solving NLP's, in section 4.7.1, is brief, since it is only intended to provide an overview of the method. An existing SQP software package was used in the implementation (Chen and Macchietto, 1989).

4.7.1 Successive Quadratic Programming (SQP)

Successive quadratic programming (SQP) is a widely discussed technique (see e.g. (Edgar and Himmelblau, 1988)) and is used here to solve the NLP problem. Basically this technique takes a local linear approximation to the constraints and a local quadratic approximation to the objective to form a quadratic program, which it solves to give a search direction s, which improves the objective. More precisely, the quadratic programming problem can be described as

$$\begin{aligned} \min \quad & s^T \nabla f(\mathrm{x}) + \tfrac{1}{2} s^T B s \\ \text{subject to} \quad & h_j(\mathrm{x}) + s^T \nabla h_j(\mathrm{x}) \quad = 0 \quad j = 1, 2, ..., meq \\ & g_j(\mathrm{x}) + s^T \nabla g_j(\mathrm{x}) \quad \geq 0 \quad j = meq + 1, ..., m \end{aligned} \tag{4.48}$$

where B is a positive definite approximation of the Hessian matrix of the Lagrangian function.

The SQP algorithm can be generally described as:

1. Let k=0. Initialise with guesses for the approximate Hessian matrix B^0, the variables (x^0), the objective ($f(\mathrm{x}^0)$), the constraints ($h(\mathrm{x}^0)$, $g(\mathrm{x}^0)$) and the gradients ($\nabla f(\mathrm{x}^0)$, $\nabla h(\mathrm{x}^0)$, $\nabla g(\mathrm{x}^0)$).

2. Form the quadratic programming subproblem (4.48) and solve for a new search direction s^k.

3. Move a step in the search direction s^k (step length given by suitable minimisation). Let k=k+1.

4. Check for convergence. If converged stop, else continue.

5. Update the values of the approximate Hessian matrix B^k, the variables (x^k), the objective ($f(\mathrm{x}^k)$), the constraints ($h(\mathrm{x}^k)$, $g(\mathrm{x}^k)$) and the gradients ($\nabla f(\mathrm{x}^k)$, $\nabla h(\mathrm{x}^k)$, $\nabla g(\mathrm{x}^k)$). Go to 2.

A more detailed discussion of this algorithm, the quadratic subproblem (4.48) and how (4.48) can be solved, is presented in (Edgar and Himmelblau, 1988), together with references to further reading on this subject.

The values of the objective function, the constraints and their gradients for any x, can be given by a differential-algebraic equation solver, which integrates the DAE's, described in (4.4), along with the sensitivity equations ($\frac{\partial x}{\partial \mathrm{X}}$ and $\frac{\partial y}{\partial \mathrm{X}}$).

4.8 Specialisation to Linear Models

The nonlinear controllability analysis presented in this chapter is highly computationally expensive, in that it uses nonlinear dynamic optimisation techniques. The technique has been specialised to linear models, to give a linear method, which is directly analogous to the nonlinear method. This linear optimal idealised control problem can be solved very efficiently using linear programming and, therefore, allows the techniques discussed in this chapter to be validated with reduced computational expense. It also provides a lower bound on the linear controllability.

The formulation of the linear problem is based on the linear relationship,

$$\delta z = P_{11}\delta w + P_{12}\delta u \tag{4.49}$$

where P_{11} and P_{12} are given by the generalised plant P. δz, δw and δu denote the deviations from the steady state of the constrained outputs, the disturbances and the control inputs respectively. The discrete elements of δz are given by

$$\delta z(k) = \sum_{l=0}^{k} P_{11}(k-l)\delta w(l) + P_{12}(k-l)\delta u(l) \tag{4.50}$$

We assume that the model used has been linearised about a steady-state given by z_{lin}, y_{lin}, w_{lin}, u_{lin}. As mentioned in section 4.4, all disturbances are considered to start from a common steady state w_0.

The nonlinear controllability problem in (4.2)

$$\min_{u_0} J(u_0) \;\; s.t. \;\; \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0$$

can be stated for the linear problem as follows

$$\begin{aligned}
\min_{u_0} J(u_0) \;\; s.t. \qquad & \forall w^p \in \bar{W} \\
& \min_{u^p \in U} c(u^p, u_0, z^p) \leq 0 \\
& z^p(k) = \sum_{l=0}^{k} P_{11}(k-l)\delta w^p(l) + P_{12}(k-l)\delta u^p(l) + z_{ref}, \;\; k = 1, N_{fh} \\
& \delta w^p(l) = w^p(l) - w_0, \;\; \delta u^p(l) = u^p(l) - u_0 \\
& z^{ref} = G_{zw}^{ss,ol}(w_0 - w_{lin}) + G_{zu}^{ss,ol}(u_0 - u_{lin}) + z_{lin} \\
& u^l \leq u^p(l) \leq u^h, \;\; l = 0, N_{fh}
\end{aligned} \tag{4.51}$$

where N_{fh} gives the length of the finite horizon over which the problem is optimised. The problem starts from a common steady state, described by w_0, which is set, and u_0, which can be an optimisation variable depending on the specific problem being solved. The disturbances, w^p, and control inputs, u^p, are discretised with the same sampling period T_{samp} as the plant elements P_{11} and P_{12}.

4.8.1 Performance Requirements

For (4.51) to be solved as an LP, both J and c must be linear functions. The constrained variable bounds in (4.32) can be included easily, since they are linear.

$$z_i^l \leq z_i^p(k) \leq z_i^h, \quad i = 1, ..., n_z \text{ for } k = 0, ..., N_{fh}$$

The objective for the problem of minimising the maximum deviation, as in (4.34), requires the maximum deviation for each disturbance ν^p to be defined linearly

$$\nu^p \geq \nu_i^p \ \ \forall i.$$

To optimise a weighted sum of the maximum deviation in each constrained output z_i, as in (4.35), we need a linear expression describing this deviation

$$\nu_i \geq \nu_i^p \ \ \forall p.$$

In both these last inequalities ν_i^p describes the maximum deviation in each constrained output, z_i^p, for each disturbance, w^p, i.e.,

$$\begin{aligned} \nu_i^p &\geq \nu_i^{p,+} \\ \nu_i^p &\geq \nu_i^{p,-} \\ -\nu_i^{p,-} &\leq z_i^p(k) \leq \nu_i^{p,+}, \text{ for } k = 0, 1, ...N_{fh} \\ \nu_i^{p,-}, \nu_i^{p,+} &\geq 0 \end{aligned}$$

The only restriction, relative to the ONDE problem, is that the economic function on the operating point $o(w_0, u_o)$ should be linear. Note that, although the model is now linear, $E(o(w)) = o(E(w)), \ E(w) = \bar{w}$ still does not hold, since the controller is nonlinear.

4.8.2 Limiting Acausal Behaviour

The technique for limiting the acausal behaviour of the idealised controller follows the same concept as for the nonlinear problem, described in section 4.5.1. Construct a set of disturbances $\bar{W}^p$ for each $w^p \in \bar{W}$, such that

$$c(\tilde{u}^{pj}, \tilde{u}_0, \tilde{w}^{pj}) \leq 0 \Rightarrow \tilde{u}_b^{pj} \neq u_b^{p,o}, \ \ \forall \tilde{w}^{pj} \in \bar{W}^p \tag{4.52}$$

and solve the optimisation problem given in (4.29).

However one element of this technique has to be altered to formulate the linear problem as an LP. The method for finding a disturbance $\tilde{w}^{\breve{p}j}$, which limits the acausal behaviour of the controller with respect to the disturbance component $w_j^{\breve{p}}$ and which would go into the set $\bar{W}^{\breve{p}}$, involves solving the following problem

$$(4.27): \qquad \tilde{J}_1(\tilde{w}_a) = \max_{\tilde{w}_a \in \tilde{W}_a} \min_{\tilde{u}^p \in \tilde{U}^{\breve{p}j}} \max_k c_k(\tilde{u}^p, u_0, \tilde{w}) \;\; s.t. \;\; \tilde{w}_b = w_b^{\breve{p}}$$

which, if $\tilde{J}_1^o > 0$, provides a candidate disturbance $\tilde{w}^o$. The problem given below

$$(4.28): \qquad \tilde{J}_2(\tilde{u}_a) = \min_{\tilde{u}_a \in U} \max_k c_k(\tilde{u}, u_0, \tilde{w}^o) \;\; s.t. \;\; \tilde{u}_b = u_b^{\breve{p},o}.$$

is then solved for $\tilde{w}^o$ and, if $\tilde{J}_2 > 0$, then we set $\tilde{w}^{\breve{p}j} = \tilde{w}^o$. The second problem, given in (4.28), can be set up as an LP, but the first, given in (4.27), cannot. We cannot maximise the maximum constraint violation for $\tilde{u}^{p,o}$ using linear programming. Therefore this problem for finding a candidate disturbance for $\tilde{w}^{\breve{p}j}$ has to be reformulated.

We want to tackle a similar problem to (4.27), but formulate it to give a problem, which is small enough to be solved by directly searching for a suitable w_a, rather than by using optimisation techniques. One way to do this is to shorten the finite horizon of the problem drastically, $l_j^{\breve{p}} \leq N_{sh} \leq N_{fh}$ where ($l_j^{\breve{p}} = t_j^{\breve{p}}/T_{samp}$), so that the set of possible step disturbances, $\tilde{W}$, is small. $\tilde{w} \in \tilde{W}$ is equivalent to $\tilde{w}$ described by $\tilde{w}_b = w_b^{\breve{p}}$ and $\tilde{w}_a \in \tilde{W}_a$, i.e.,

$$\tilde{w} \in \tilde{W} \Rightarrow \tilde{w}_h(l) = \begin{cases} w_h^{\breve{p}}(lT_{samp}), & l = 0, 1, ...l_h - 1 \\ \left.\begin{array}{ll} w_h^{\breve{p}}(lT_{samp}), & \text{if } t_h^{\breve{p}} < t_j^{\breve{p}} \\ sign_h, & \text{if } t_h^{\breve{p}} \geq t_j^{\breve{p}} \end{array}\right\} & l = l_h, l_h + 1, ... \end{cases} \tag{4.53}$$

where $sign_h = w_h^h$ or w_h^l, since the disturbance can step either to its upper bound, w_h^h, or its lower bound, w_h^l, and l_h is a discrete sample time, between $l_j^{\breve{p}} \leq l_h \leq N_{sh}$, when $\tilde{w}_h$ steps to $sign_h$. The different choices of $sign_h$ and t_h create the finite set $\tilde{W}$ of $(2(N_{sh} - l_j^{\breve{p}} + 1))^m$ step disturbances, where m is the number of components $w_h^{\breve{p}}$, $1 \leq h \leq n_w$, for which $t_h^{\breve{p}} \geq t_j^{\breve{p}}$.

Note that $l_j^{\check{p}} = t_j^{\check{p}}/T_{samp}$ where $t_j^{\check{p}}$ is the time at which a step occurs in $w_j^{\check{p}}$, given in (4.18), and the sampling time, T_{samp}, is selected to be a factor of $t_j^{\check{p}}$. $t_h^{\check{p}}$ is the time at which a step occurs in the hth component $w_h^{\check{p}}$ of the disturbance $w^{\check{p}}$.

As long as N_{sh} is chosen small enough then for every $\tilde{w} \in \tilde{W}$ and every $\tilde{u}^p \in \tilde{U}^{\check{p}j}$, $(p = 1, .., n_{dist})$, the constraints at the lth sample time

$$c(\tilde{u}^p(l), u_0, \tilde{w}(l)) \quad l = 0, 1, ..., N_{sh}$$

can be calculated without excessive computational expense. If for some $\tilde{w} \in \tilde{W}$ it is found that

$$\max_k \max_l c_k(\tilde{u}^p(l), u_0, \tilde{w}(l)) > 0 \;\; \forall \tilde{u}^p \in \tilde{U}^{\check{p}j}$$

then this disturbance is a candidate disturbance $\tilde{w}^c$. There might be more than one such disturbance found in $\tilde{W}$, in which case this gives a set of candidate disturbances $\tilde{w}^c \in \tilde{W}^c$. On the other hand, no such disturbance might be found in $\tilde{W}$. In this case the shortened finite horizon N_{sh} should be increased and the process repeated until either at least one $\tilde{w}^c$ has been found or $N_{sh} = N_{fh}$.

This technique, therefore, can be used to provide a set $\tilde{W}^c$ of feasible candidate disturbances, rather than the optimal candidate disturbance $\tilde{w}^o$ produced by the optimisation problem (4.27). To select one optimal candidate disturbance from $\tilde{W}^c$ find the $\tilde{w}^o \in \tilde{W}_c$ which optimises,

$$\max_{\tilde{w}^c \in \tilde{W}^c} \min_{\tilde{u}^p \in \tilde{U}^{\check{p}j}} \max_k \left[\frac{\max_l c_k(\tilde{u}^p(l), u_0, \tilde{w}^c(l))}{l_k^{max}} \right]$$

where l_k^{max} is the time of the maximum violation in constraint c_k, i.e.,

$$\max_l c_k(\tilde{u}^p(l), u_0, \tilde{w}^c(l)) = c_k(\tilde{u}^p(l_k^{max}), u_0, \tilde{w}^c(l_k^{max})).$$

This selects a candidate disturbance, which provides a violation which is both large and fast. These two attributes should make it hard for problem (4.28) to find any u_a which ensures feasible behaviour.

Now that an optimal candidate disturbance, $\tilde{w}^o$, has been found this can be passed to the optimisation problem (4.28), exactly as for the nonlinear version, and, if $\tilde{J}_2 > 0$, then we set $\tilde{w}^{\breve{p}j} = \tilde{w}^o$ and $\tilde{w}^{\breve{p}j} \in \tilde{W}^{\breve{p}}$. This procedure is repeated for every component $w_j^{\breve{p}}$, $1 \leq j \leq n_w$, in the disturbance $w^{\breve{p}} \in \bar{W}$ for which

$$t_j^{\breve{p}} > \min_l t_l^{\breve{p}},$$

In this way the sets $\bar{W}^p$ for each $w^p \in \bar{W}$ can be constructed and the optimisation problem in (4.29) solved as an LP.

However as mentioned in section 4.5.1 this technique may not find any appropriate $\tilde{w}^{pj}$'s to go into the set $\bar{W}^p$ and, even when it can, it does not guarantee causality.

This linear problem allows us to solve many of the same problems that might be applied to the nonlinear model with this computationally cheaper linear alternative. Due to the linear nature of this problem it can be implemented in an extremely efficient manner. An example is presented illustrating the application of this linear problem with the technique for limiting acausal behaviour.

4.8.3 An Example of Limited Acausal Behaviour

The linear method and the technique for limiting acausal behaviour is demonstrated by applying it to a simple linear model given by

$$z = \begin{bmatrix} \frac{0.5(s+1)}{(s^2+s+1)} & \frac{0.5(s-1)e^{-4s}}{(s^2+0.8s+1)} & \frac{1.5e^{-4s}}{(1.5s+1)} \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ u \end{pmatrix} \tag{4.54}$$

which is then discretised using the Tustin method for a sample period of 0.5 seconds. All the tests are run for a finite horizon of 70 samples.

In the following, it is assumed that feedforward control can be used and, therefore, a realisable controller would be able to respond as soon as the disturbance had occurred. This means that time delays between the disturbances and the measurements need not be considered.

Test 1

The open-loop step responses, for positive unit steps, are shown in Figure 4.2 for the first disturbance and the second disturbance. The greatest peak for w_1 is 0.6480 at

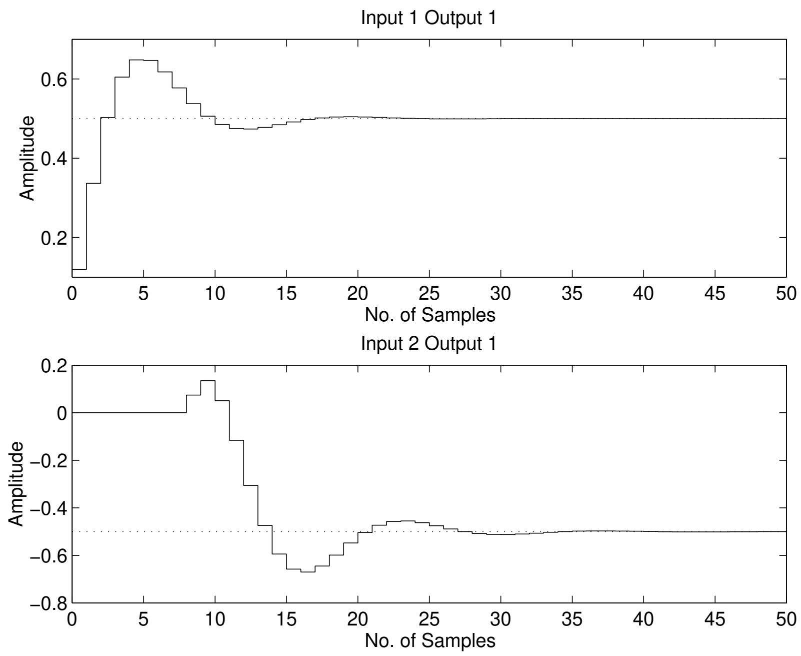

Figure 4.2: The step responses of disturbances w_1 and w_2

sample 4 and for w_2 it is -0.6696 at sample 16. This suggest that the worst open-loop disturbance might be given by the disturbance w^{OL} in Figure 4.3. which gives a peak at sample 16 of 1.3176 as shown in Figure 4.4.

So making the objective to minimise the maximum deviation in z, we optimise the problem

$$\min_u \max_k |z(k)| \ \ s.t. -1 \leq u \leq 1 \ \ \forall w \in \bar{W} \tag{4.55}$$

where the set of worst step disturbances is given as $\bar{W} = w^{OL}$ so $n_{dist} = 1$. The optimal result for this problem is 0.1129. The controller selected for this optimum will utilise the first 12 samples to not only handle the step in w_2, but to prepare for the future step in w_1. As can be seen in Figure 4.5, the controller selects to be on the lower bound for much of the time. How much of this behaviour will be due to it coping with the existing step in w_2 and how much can be attributed to its future knowledge of w_1 is unclear.

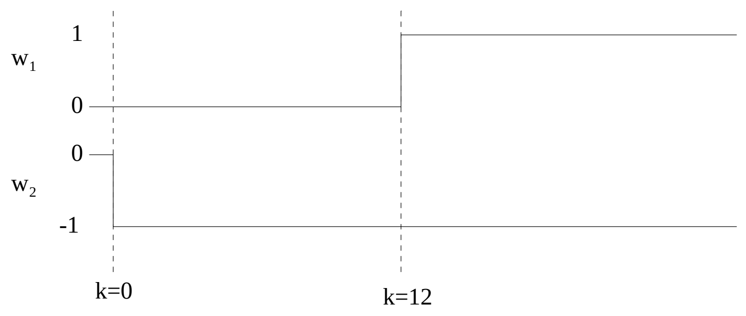

Figure 4.3: The worst open-loop step disturbance w^{OL}

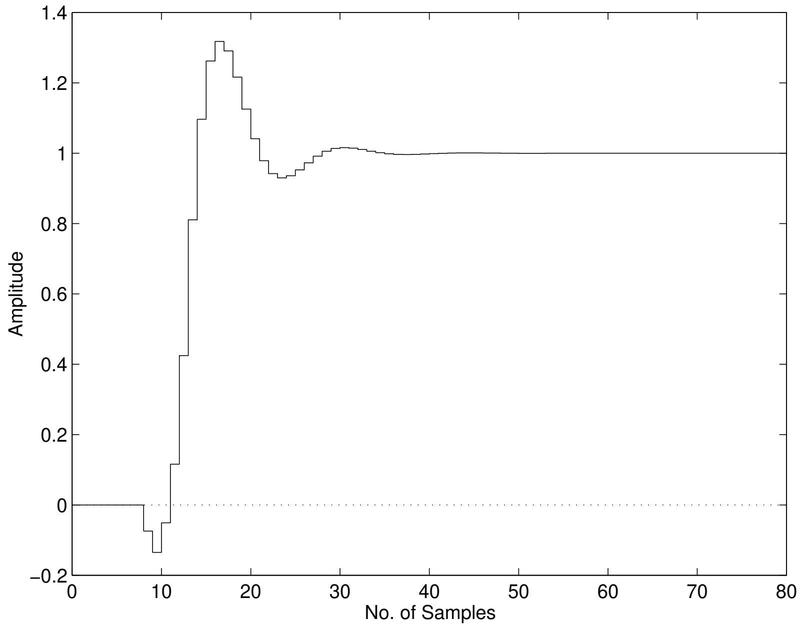

Figure 4.4: The worst open-loop response

If we held the control until both steps have occurred then there is no chance for the controller to be acausal. This involves solving

$$\min_u \max_k |z(k)| \;\; s.t. \;\; \begin{array}{c} -1 \leq u \leq 1 \\ u(k) = 0, \quad k = 0, 1, ...11 \end{array} \qquad \text{for } w^{OL} \tag{4.56}$$

which gives an optimal value of 1.3176. This result is much worse than that for the acausal controller, since the controller is not allowed to do anything about the step in w_2 until after the occurrence of the step in w_1. In fact, this optimal result means, that the controller has not been able to reduce the open-loop peak due to the worst disturbance at all. Therefore this performance is probably much worse than it could be for a real controller. The performance of an optimal causal controller would be expected to be somewhere in between these two values.

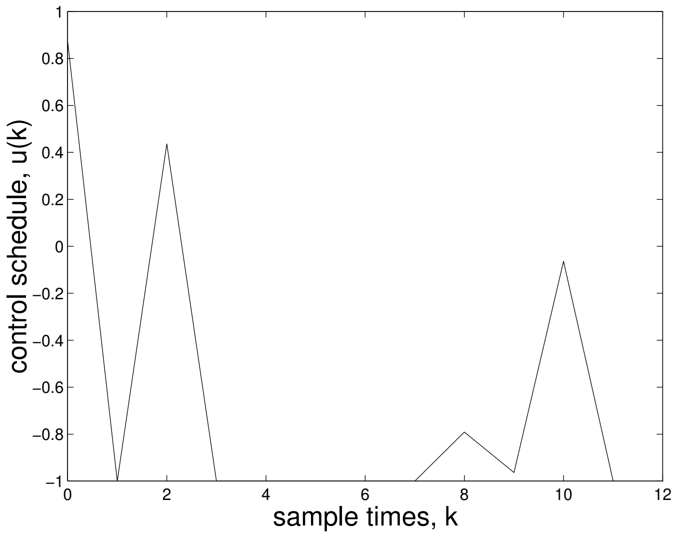

Figure 4.5: The acausal control schedule for $k < 12$

Therefore using the technique presented earlier for limiting the acausal nature of u we find a disturbance $\tilde{w}^{OL}$ for which

$$(\min_{-1\le\tilde{u}\le 1}\max_{k}|\tilde{z}(k)|) > 0.1129 \quad \text{where} \quad \begin{cases} \tilde{z} = P_{11}\tilde{w}^{OL} + P_{12}\tilde{u} \\ \tilde{u}(k) = u^{o}(k), \quad k = 0, 1, ..., 11 \end{cases} \tag{4.57}$$

and

$$\tilde{w}^{OL}(k) = w^{OL}(k), \quad k = 0, 1, ..., 11 \tag{4.58}$$

where u^o is the optimal controller from the first optimisation, i.e., the acausal controller. The disturbance selected is shown in Figure 4.6.

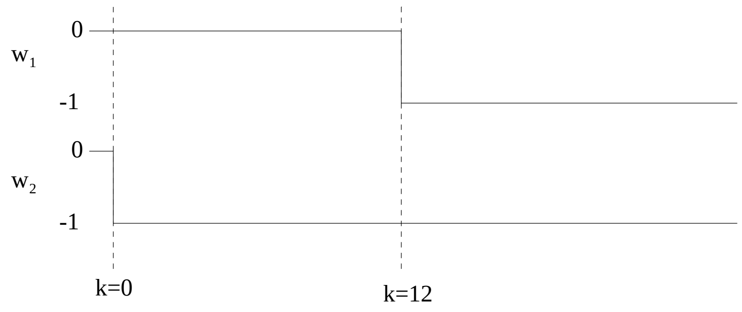

Figure 4.6: The disturbance selected to limit acausal behaviour $\tilde{w}^{OL}$

The optimal result of the problem

$$\begin{aligned} \min_{u,\tilde{u}} \max(\max_k |z(k)|, \max_k |\tilde{z}(k)|) \qquad & s.t. \\ & -1 \leq u \leq 1 \\ & -1 \leq \tilde{u} \leq 1 \\ & \tilde{u}(k) = u(k), \quad k = 0, 1, ..., 11 \\ & \text{for } \tilde{z} = P_{11}\tilde{w}^{OL} + P_{12}\tilde{u} \\ & \text{and } z = P_{11}w^{OL} + P_{12}u \end{aligned} \tag{4.59}$$

is 0.6480 which as expected sits between the acausal optimal result, 0.1129, and the optimal result for the controller held still for the initial 12 samples, 1.3176. The control schedule selected for the first 12 samples is shown in Figure 4.7. As might be expected less of the controllers time is spent on its lower bound preparing for the future appearance of w_1. This optimal result, 0.6480, means that the optimal controller has removed the peak due to w_2 using feedforward control, but cannot reduce the peak due to w_1. This suggests that the behaviour is most limited by the input constraints and the non-minimum phase characteristics between u and z. This technique cannot guarantee to always remove all of the acausal element of the controller, but the result certainly provides a tighter bound than that given by the original optimisation problem.

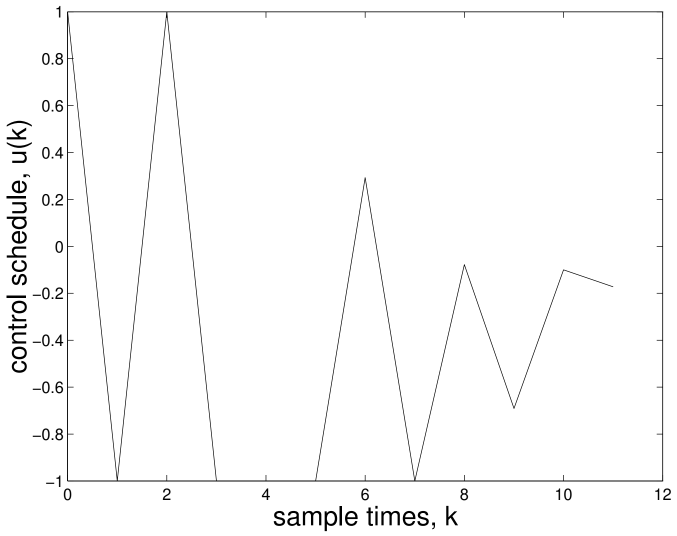

Figure 4.7: The control schedule with limited acausality for $k < 12$

Test 2

An interesting test, to show that this technique can limit the acausal nature of the controller, is to use it on a problem where the disturbances are all delayed by the same, k_{del}, samples. In this case the result for a causal controller can be found by just holding the controller still for these first k_{del} samples or simply by applying the disturbance with no delay.

Therefore select a disturbance, such as w^{del} in Figure 4.8, and solve problem (4.55) for it. The optimal result this provides for the acausal controller is 0.1224. The

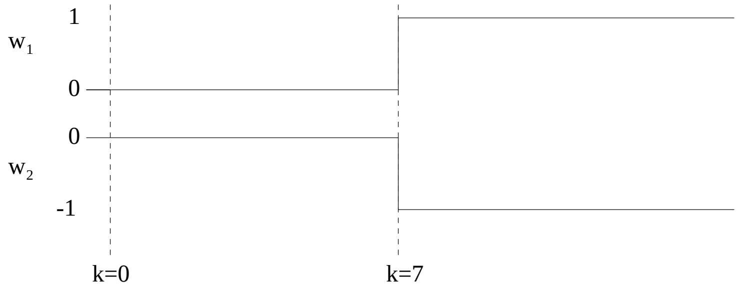

Figure 4.8: Disturbance w^{del}, both w_1 and w_2 delayed by $k = 7$

disturbance found to satisfy

$$(\min_{-1\leq\tilde{u}\leq 1}\max_k |\tilde{z}(k)|) > 0.1224 \quad \text{where} \quad \begin{cases} \tilde{z} = P_{11}\tilde{w}^{del} + P_{12}\tilde{u} \\ \tilde{u}(k) = u^o(k), \quad k = 0, 1, ..., 6 \end{cases} \tag{4.60}$$

is shown in Figure 4.9. The technique does not change w_2 to limit the acausal behaviour of the controller, since the controller uses feedforward to handle this disturbance, not acausal knowledge.

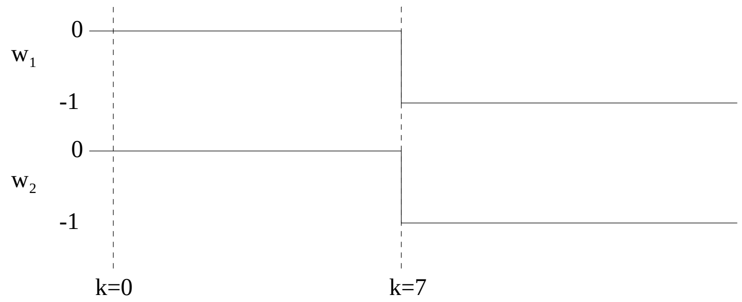

Figure 4.9: The disturbance selected to limit acausal behaviour $\tilde{w}^{del}$

The optimal result with limited acausal behaviour, given by problem (4.61), is 0.6480.

$$\begin{aligned} \min_{u,\tilde{u}} \max(\max_k |z(k)|, \max_k |\tilde{z}(k)|) \qquad & s.t. \\ & -1 \leq u \leq 1 \\ & -1 \leq \tilde{u} \leq 1 \\ & \tilde{u}(k) = u(k), \quad k = 0, 1, ..., 7 \\ & \text{for } \tilde{z} = P_{11}\tilde{w}^{del} + P_{12}\tilde{u} \\ & \text{and } z = P_{11}w^{del} + P_{12}u \end{aligned} \tag{4.61}$$

The optimal result for a causal controller, i.e., when w^{del} is applied with no delay as in Figure 4.10, was found to also be 0.6480, showing that in this particular case the technique removes all acausal elements from the controller. As for the previous test, the results of this test suggest, that the achievable behaviour for this problem is most limited by the input constraints and the non-minimum phase characteristics between u and z.

Figure 4.10: Disturbance w^{del} with no delay

4.9 Implementation

The software was programmed in FORTRAN and used two subroutines, DASOLV (Jarvis and Pantelides, 1992) and SRQPD (Chen and Macchietto, 1989), which are existing software packages. Its general layout is shown in Figure 4.11.

A differential-algebraic equation solver, DASOLV, was used to solve the system DAE's

$$\left.\begin{array}{l} F(\dot{x}^p, x^p, y^p, u^p, w^p, t) = 0 \text{ for } u^p \in \bar{U},\ w^p \in \bar{W} \\ \dot{x^p}(t_0) = 0 \end{array}\right\} \tag{4.62}$$

as well as the to provide the sensitivities

$$\frac{\partial x^p}{\partial \mathrm{x}},\ \frac{\partial y^p}{\partial \mathrm{x}},\ \frac{\partial u^p}{\partial \mathrm{x}},\ \frac{\partial w^p}{\partial \mathrm{x}}. \tag{4.63}$$

The residuals describing the function F, as well as the systems jacobian matrix describing $\frac{\partial F}{\partial x^p}$, $\frac{\partial F}{\partial y^p}$, $\frac{\partial F}{\partial u^p}$ and $\frac{\partial F}{\partial w^p}$, must be supplied to DASOLV through FORTRAN subroutines.

These last two FORTRAN subroutines can be automatically generated, if the non-linear model is programmed in gPROMS and coded as if to go on to solve a gOPT problem. Otherwise these could be entered by hand into FORTRAN subroutines to be

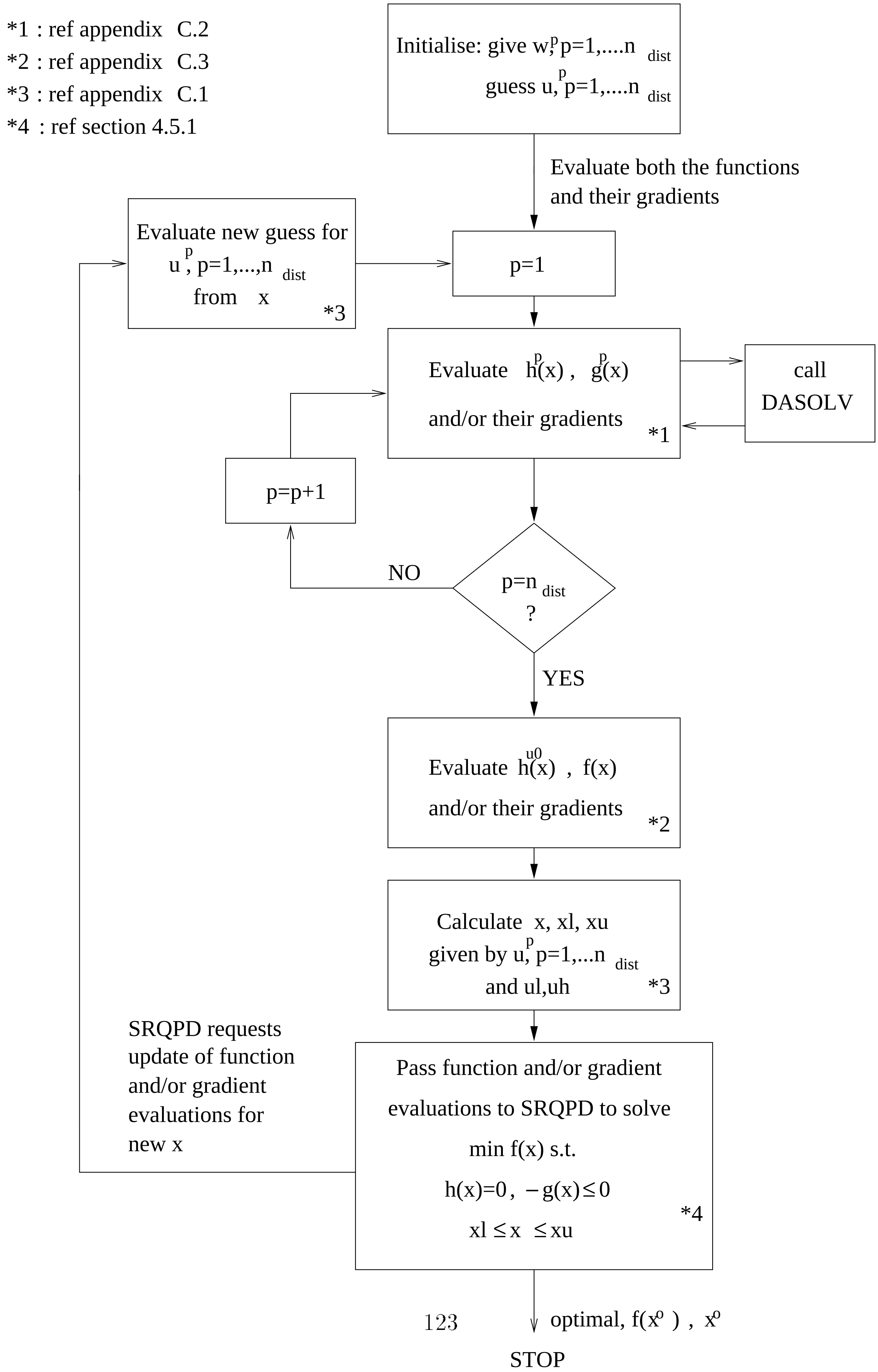

Figure 4.11: Software for the nonlinear controllability analysis

called by DASOLV, although this might be rather time consuming depending on the size of the problem.

The values evaluated by DASOLV are used to set up the NLP, as described in section 4.7

$$
\begin{aligned}
\min_{\mathrm{x}} f(\mathrm{x}) & & \\
\text{subject to } h_j(\mathrm{x}) &= 0 \quad j = 1, 2, ..., meq \\
g_j(\mathrm{x}) &\geq 0 \quad j = meq + 1, ..., m \\
\mathrm{x}_l \leq & \ \mathrm{x} \ \leq \mathrm{x}_u
\end{aligned}
$$

This is passed to a package (SRQPD) which implements a SQP method, to solve 4.64 to find a local minimum. SRQPD follows the type of SQP algorithm described in section 4.7.1. It requires initial guesses of the optimisation variables x^0, the functions $f(\mathrm{x}^0)$, $h(\mathrm{x}^0)$ and $g(\mathrm{x}^0)$, and the gradients $\nabla f(\mathrm{x}^0)$, $\nabla h(\mathrm{x}^0)$ and $\nabla g(\mathrm{x}^0)$. There is no need to provide a guess of the Hessian matrix B^0, since SRQPD will automatically generate this. The routine will automatically request further function and/or gradient evaluation for new values of x, until it has converged.

4.10 Review

A nonlinear controllability analysis technique, complementary to the linear technique presented in the previous chapter, has been developed. It is based on the optimal idealised control problem

$$\min_{u_0} J(u_0) \ \ s.t. \ \ \max_{w^p \in \bar{W}} \min_{u^p \in U} c(u^p, u_0, w^p) \leq 0 \tag{4.64}$$

This technique provides a lower (optimistic) bound on the controllability and has the following properties:

- It allows constraints and disturbances to be defined in the time domain. A wide range of typical process performance requirements can be captured, including any expressed as nonlinear functions, unlike the linear technique. However the disturbance description is more restricted than for the linear case, i.e., a finite set of step disturbances.

- It optimises the achievable performance for an idealised controller, providing an optimistic bound on the controllability. A technique for tightening this optimistic bound, by limiting the acausal nature of the idealised controller, has been developed.

- It incorporates many of the fundamental limitations on controllability. The limitation due to NMP characteristics, appearing between the control inputs and the regulated outputs, is captured directly by the DAE's of the system. Limitation due to such characteristics appearing between the disturbances and the measured variables is not included. However, a suggestion for incorporating the limitation due to time delays between w and y is described in section 4.5.1. The input constraints can be incorporated as magnitude bounds on the optimisation variables. Any limitation due to measurement noise can be captured by including the noise in the disturbance vector and altering the DAE's describing the system, so that these (noise) disturbances are added on to the appropriate measurements. The only fundamental limitation completely ignored in this technique is uncertainty.

- It is formulated and solved as an NLP and, therefore, is computationally expensive. However a linear alternative which can be solved as an LP has been developed.

Some particular shortcomings of this technique is the degree of idealisation of the controller which makes the tightness of this lower bound unclear, the disturbance set is finite and user specified, and the technique is computationally expensive. However we have attempted to tackle most of these criticisms. A technique for limiting the acausal nature of the controller has been developed, to tighten the lower bound. The linear technique discussed, in the previous chapter, can be used to give good estimates of the worst process step disturbances, if the process disturbances for the plant have not been identified. Finally, we have developed a specialisation of this nonlinear technique for linear models, which can be solved efficiently as an LP. This allows the nonlinear techniques to be tested for a much reduced computational cost and provides a lower bound on the linear controllability. It also allows the correlation of the linear and nonlinear models, for these controllability tests, to be checked.

This method answers many of the requirements set out in section 2.4.1 and provides a strong result. If the process fails this test, then it means, either the nominal process will not be able to meet the performance specifications with any controller, or the NLP solver has failed to find a global optimum.

Conclusions on both the linear and nonlinear techniques are discussed in the final chapter of this thesis.

Chapter 5

Case Studies

In this chapter several case studies are presented to illustrate the application of the controllability techniques developed. The first was for a X29 aircraft and was based on a problem described in "Control of Uncertain Systems: A Linear Programming Approach" (Dahleh and Diaz-Bobillo, 1995). The results on this example are checked against this published example. This example is not an illustration of controllability analysis *per se*, but is used to validate the software and explore some formulation issues. The second example is an exothermic plug flow reactor and is based on an industrial process system example presented in "Controllability analysis and modelling requirements: an industrial example" (Walsh and Malik, 1995). This tests the analysis for a system with time delays and input constraints. The results of the linear performance optimisation problem, for both persistent disturbances and step disturbances, are discussed. The third case study is an industrial problem reactor control problem that was presented in "A case study in control structure selection for a chemical reactor" (Walsh *et al.*, 1997). The OLDE analysis is applied to the linearised model of this problem and the results compared to some nonlinear controllability analysis. Finally a model of an evaporator, presented in detail in "Applied Process Control: A Case Study" (Newell and Lee, 1989), is subjected to the OLDE analysis, the ONDE analysis and its linear specialisation, and the results for the linear and nonlinear analysis compared.

5.1 Case Study 1: Pitch Axis Control of the X29 Aircraft

A design example was selected from "Control of Uncertain Systems: A Linear Programming Approach" (Dahleh and Diaz-Bobillo, 1995) to set up the solution technique. The X29 aircraft is statically unstable due to its forward swept wing design, which places the centre of gravity behind the aerodynamic centre of pressure. Therefore this design example is to design a digital pitch axis controller for the X29 aircraft. The three control surfaces, canard wings, flaperons on the main wings and strakes on the tail, were lumped together, for simplicity, giving one actuator with first order dynamics. Also the gyroscopes and accelerometers were modelled by a sensor with negligible dynamics. This gives the continuous-time SISO plant given below,

$$\hat{P}(s) = \underbrace{\frac{(s+3)}{(s+10)(s-6)}}_{airframe} \underbrace{\frac{20}{(s+20)}}_{equiv.actuator} \underbrace{\frac{(s-26)}{(s+26)}}_{overhead}. \tag{5.1}$$

which has one NMP zero and one unstable pole

Dahleh and Diaz-Bobillo setup the ℓ_1 performance objectives to minimise the effect of the disturbance, w, on both the weighted control sequence, z_1, (the controller effort) and the weighted output, z_2, (Figure 5.1). Therefore the ℓ_1 problem is posed as follows

$$\nu_o = \inf_{K stab.} \left\| \begin{matrix} W_1 K S \\ W_2 S \end{matrix} \right\|_1 \tag{5.2}$$

where $S = (I - PK)^{-1}$=sensitivity function.

The weights were chosen in the book, not only to accommodate the trade-offs between low frequency disturbance rejection and controller effort, but also to emphasise the frequency regions which corresponded to the spectral content of the exogenous disturbance.

$$\hat{W}_1(s) = 0.01 \qquad \hat{W}_2(s) = \frac{(s+1)}{(s+0.001)} \tag{5.3}$$

where $\hat{W}_2$ emphasises disturbance rejection at low frequencies

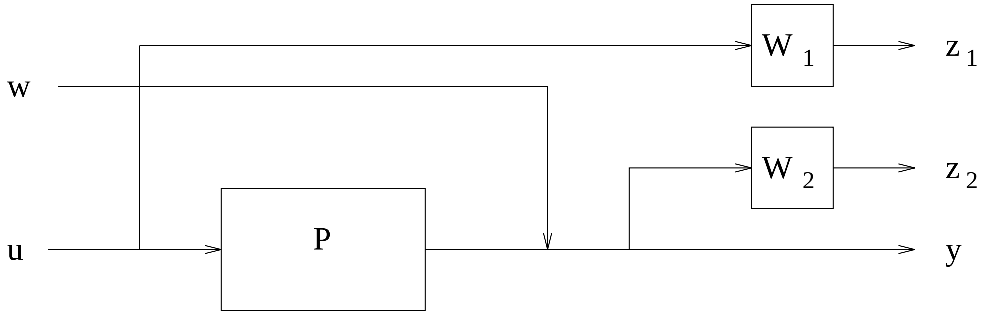

Figure 5.1: The X29 problem in standard form

The continuous system is represented as

$$\begin{pmatrix} z_1 \\ z_2 \\ y \end{pmatrix} = \begin{pmatrix} 0 & W_1 \\ W_2 & W_2 P \\ I & P \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix} = \frac{1}{w_{2d} d} \begin{pmatrix} 0 & w_1 w_{2d} d \\ w_{2n} d & w_{2n} n \\ w_{2d} d & n w_{2d} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix} \tag{5.4}$$

This was discretised using a zero order hold with sampling period $T_s = 1/30$, as used in the book.

Since the system has been represented by an equivalent SISO plant, then the number of measured outputs n_y, control inputs n_u and exogenous inputs (disturbances) n_w are all 1, whilst the number of regulated outputs n_z is 2. Therefore $n_w = n_y$ and $n_z > n_u$, giving a two-block column problem. This is a multiblock problem, also called bad rank or nonsquare, therefore both zero and rank interpolation conditions are required, which suggests the use the Delay Augmentation method to avoid the infinite constraints given by the rank interpolation conditions.

This problem is solved in Dahleh and Diaz-Bobillo (1995) for values of the DA delay ,N, up to 80. They give the lower bound at $N = 80$ as 4.054. The software developed in this thesis gives the convergence shown in Figure 5.2 with a value of the lower bound at $N = 80$ of $\underline{\eta}_N = 4.054$ (the lower bound, $\underline{\eta}_N$, and the upper bound, $\overline{\eta}_N$, are discussed in Appendix A.1). The LP was solved with several different LP solvers (based on MINOS, CPLEX and OSL subroutines) which all gave the same results.

Also an upper bound is calculated, as discussed in Appendix A.1, by extracting the optimal Q_N° at $N = 80$ from the closed loop response Φ_N° and using the upper $n_u \times n_y$,

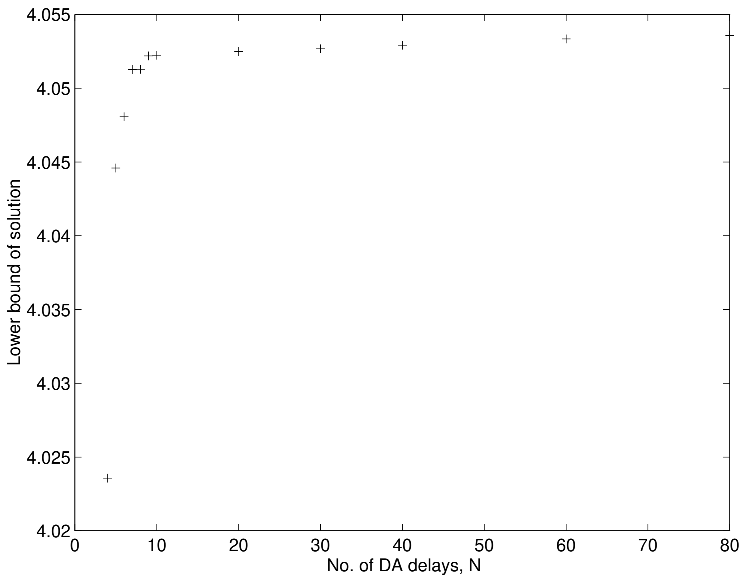

Figure 5.2: Convergence of DA methods for X29 for zero order hold

i.e., 1×1, matrix which is Q^0_{11} to find

$$\|\bar{\Phi}^o\|_1 = \|H - UQ^o_{11}V\|_1 = \sum_{k=0}^{\infty} |\phi(k)|$$

This is done by summing the absolute values of the elements of the discrete impulse response over a long, but finite, horizon, i.e.,

$$\overline{\eta}_N = \max_i \sum_{j=1}^{n_w} \sum_{k=0}^{k_0} |\bar{\phi}_{ij}(k)|$$

where k_0 is a finite number, which is chosen to be large enough that the elements of $\bar{\phi}(k)$ are close to 0. It was found, that for k's greater than 500 $\max_i |\bar{\phi}_i(k)| < 10^{-9}$, (in this example $n_w = 1$ and $n_z = 2$ so let $\bar{\phi}_{ij} = \bar{\phi}_i$ for $i = 1, 2$) . We allowed $k_0 = 3000$ and found that

$$\sum_{k=0}^{3000} |\phi(\bar{k})| = \begin{bmatrix} 4.054 \\ 4.084 \end{bmatrix}$$

This gives an upper bound of $\overline{\eta}_N = 4.084$, therefore

$$4.054 \leq \nu_o \leq 4.084.$$

Thus we have converged the lower bound to within 1% of the optimal value ν_o, since $100\% \times \left(\frac{\overline{\eta}_N - \underline{\eta}_N}{\underline{\eta}_N}\right) = 0.74\%$. These results compare well with the published results.

It is interesting to note that, if a bilinear transform, ie, the Tustin's transform $s = \frac{2(z-1)}{T(z+1)}$, were used to discretise the system, rather than zero order hold, as in this problem, a better solution, ie, a lower ν_o, might be expected. This is because by using zero order hold we have restricted the group of digital controllers that the objective can be minimised over. Similarly an improvement in the norm could be found if the sampling period for the zero order hold were decreased. To use this technique for controllability analysis we want the value of ν_o to be as unbiased by the discretisation as possible, therefore we would want to use as good an approximation for the discretisation as possible, such as the Tustin transform.

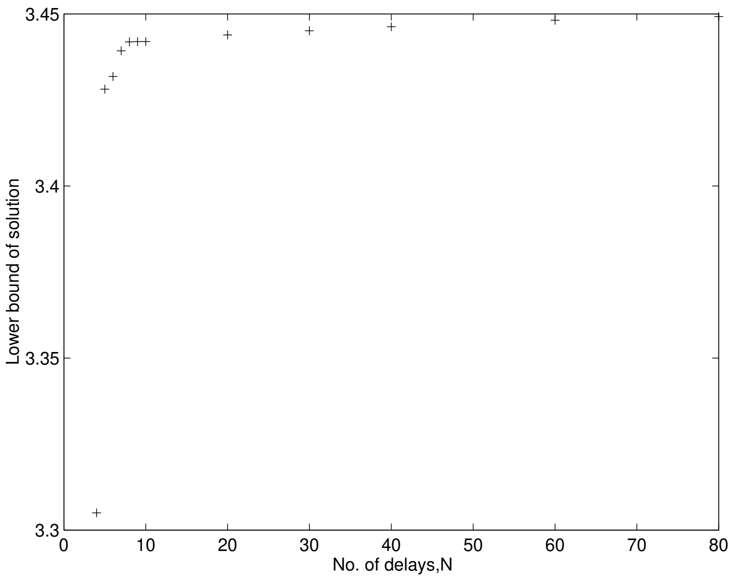

Figure 5.3: Convergence of DA methods for X29 for Tustin

The results using the Tustin transform are shown in Figure 5.3. The ν_o has been reduced as expected.

5.1.1 Conclusions

The software developed in this project has been validated against a published example in Dahleh and Diaz-Bobillo (1995) and the calculation of an upper bound demonstrated.

Also the importance of the discretisation method adopted has been highlighted. It is well known that zero order hold sampling degrades the achievable performance by an amount comparable to a delay of half the sampling interval. The use of this approximation will, therefore, require quite small sampling intervals to approach the continuous case closely. The zero order hold approximation would be appropriate only if the sampling interval and zero order hold interpolation was specified as part of the problem definition, e.g. due to the use of a particular computer control system. The Tustin approximation is designed to give a closer match between the discrete time model and the original continuous time model.

5.2 Case Study 2: An Industrial Exothermic Plug Flow Reactor Example

The simple model used in "Controllability analysis and modelling requirements: an industrial example" (Walsh and Malik, 1995) was obtained by constructing a transfer function matrix of gain-delay-lag elements, ie, $\frac{ke^{-t_d s}}{1+\tau s}$, using step response tests. This exothermic plug flow reactor is shown in Figure 5.4.

The performance requirements are given as a set of target steady-state values and constraints for the variables T_{in} and T_{out}:

variable	steady-state value	upper limit	lower limit
T_{in}	400 $°C$	none	300 $°C$
T_{out}	600 $°C$	650 $°C$	none

Note that the desired closed loop response is otherwise unspecified. There are no steady-state degrees of freedom, so u_0 is fixed. In a previous control study serious difficulties had been encountered in meeting the constraints above, even assuming rate limits on the disturbances. The controllability analysis below, therefore, focuses on feasibility and is also presented in Chenery and Walsh (1997).

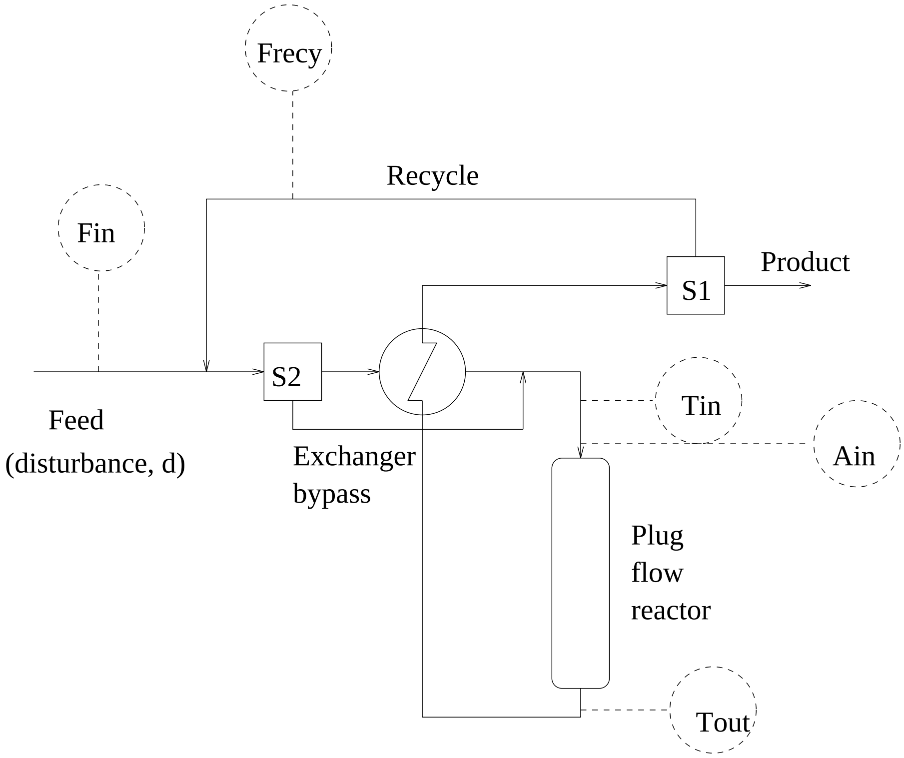

Figure 5.4: The exothermic plug flow reactor (PFR)

There are only two possible manipulated variables, S_1, the split between the product and the recycle and, S_2, the split between the heat exchanger and the bypass ($S_1 = 1$ means 100% recycled and $S_2 = 1$ means 100% bypassed). F_{in} is determined by the requirements of another plant. There are four measured variables, A_{in}, a delayed composition measurement of the key reactant, $F_{recycle}$, T_{in} and T_{out}. The disturbance, d, is a well defined change between two operating regimes,A and B, which involves three feed component flowrates. Since the disturbance only occurs in a single direction with respect to these flowrates, it is represented as a single disturbance. The temperature measurements, disturbance and manipulated variables are all normalised by their maximum acceptable deviation. The transfer function matrix obtained is shown below (with time in hours):

$$\begin{pmatrix} A_{in} \\ F_{recycle} \\ T_{out} \\ T_{in} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{2.4e^{-.008s}}{1+.02s} & \frac{-1e^{-.008s}}{1+.02s} & 0 \\ .13 & 3.5 & 0 \\ \frac{16.6e^{-.25s}}{1+.6s} & \frac{-7.4e^{-.35s}}{1+.6s} & \frac{-4.8e^{-.45s}}{1+.6s} \\ \frac{4.3e^{-.25s}}{1+.6s} & \frac{-1.9e^{-.35s}}{1+.6s} & -1.8 - \frac{.5e^{-.4s}}{1+.6s} \end{pmatrix}}_{T} \begin{pmatrix} d \\ S_1 \\ S_2 \end{pmatrix}. \tag{5.5}$$

Note: There was a pronounced oscillatory response to step changes, which made finding the gain-delay-lag elements by step response testing difficult. The model was fitted to the mean value of the oscillation and the delay chosen to encompass the initial "inverse response".

The measured outputs, y, and regulated outputs, z, are given by

$$y = \begin{pmatrix} A_{in} \\ F_{recycle} \\ T_{out} \\ T_{in} \end{pmatrix}, z = \begin{pmatrix} T_{out} \\ T_{in} \end{pmatrix} \tag{5.6}$$

whilst the exogenous input, w, and the control inputs, u, are given by

$$w = d, u = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}. \tag{5.7}$$

The disturbance, w, is asymmetric, ie, $0 \leq w \leq 1$, therefore the system should be rescaled so that the description of all disturbances as $\|w\|_\infty \leq 1$ is tight. This gives the smallest ℓ_1 set encompassing the actual disturbance and, therefore, minimises conservatism in the analysis. This requires defining the new reference disturbance as the mean 0.5, ie, let $\tilde{w} = 2(w - 0.5)$ so that $-1 \leq \tilde{w} \leq 1$. P becomes $\tilde{P} = PS$ where the rescaling matrix S is given by,

$$S = \begin{pmatrix} S_w & \mathbf{0} \\ \mathbf{0} & S_u \end{pmatrix} \tag{5.8}$$

$S_w = 0.5$ and $S_u = \min(u_h - u_{ref}, u_{ref} - u_l)$ for $u_{ref} = -K_{12}^{-1}K_{11}/2$, ($K_{11}$ and K_{12} are the steady state gain matrices of $P_{11}$and P_{12} respectively), $u_h = -u_l = (3\ \ 1)^T$. This means that the problem is being scaled such that $-3 \leq S_1 \leq 3$ and $-1 \leq S_2 \leq 1$. S_1 is scaled to lie between -3 and 3, as discussed in Walsh and Malik (1995), to account for observed mismatch between linear and nonlinear steady-state behaviour associated with the actuators.

The rescaled transfer function matrix relating the scaled variables $(\tilde{d}\ \tilde{S}_1\ \tilde{S}_2)^T$ to $(A_{in}\ F_{recycle}\ \tilde{T}_{out}\ \tilde{T}_{in})^T$ is given by

$$T = \begin{pmatrix} \frac{1.2e^{-.008s}}{1+.02s} & \frac{-1.89e^{-.008s}}{1+.02s} & 0 \\ .065 & 6.61 & 0 \\ \frac{8.3e^{-.25s}}{1+.6s} & \frac{-13.98e^{-.35s}}{1+.6s} & \frac{-4.71e^{-.45s}}{1+.6s} \\ \frac{2.15e^{-.25s}}{1+.6s} & \frac{-3.59e^{-.35s}}{1+.6s} & -1.77 - \frac{.49e^{-.4s}}{1+.6s} \end{pmatrix}$$

A state space realisation of the transfer function was found and discretised using the bilinear Tustin method.

To discretise this model the sampling period is chosen to be the smallest time delay, ie, $T_s = 0.008$. The fact that some time delays in the model are as much as $\frac{.45}{T_s} = 56.25$ times larger than T_s means that the discretised transfer function matrix will have some very high order terms. This, in turn, means that the state space representation of the system will have many states, which will make the solution method slow. Therefore each gain-delay-lag element, for which $ceil(t_d/T_s) > 1$, was approximated as follows

$$T_{ij}(s) = k_1 + Td_n(s)\frac{k_2}{1+\tau s} \tag{5.9}$$

where $Td_n(s)$ is the n^{th} order Pade approximation of the time delay $e^{-t_d s}$ ($n = 4$ was used).

Note: The n^{th} order Pade approximation of a time delay t_d is given by

$$e^{-t_d s} \approx \frac{1 + \sum_{i=0}^{n} \frac{(-t_d s)^i}{i!}}{1 + \sum_{i=0}^{n} \frac{(t_d s)^i}{i!}} \tag{5.10}$$

since

$$\left[e^{-t_d s} = (1+e^{-t_d s})\sum_{i=0}^{\infty}(-e^{t_d s})^i = \frac{1+e^{-t_d s}}{1+e^{t_d s}} \text{ and } e^{t_d s} = \sum_{i=0}^{\infty}\frac{(t_d s)^i}{i!}s\right]$$

The problem was initially set as follows.

$$y = \begin{pmatrix} A_{in} \\ F_{recycle} \\ \tilde{T}_{out} \\ \tilde{T}_{in} \end{pmatrix}, z = \begin{pmatrix} \tilde{T}_{out} \\ \tilde{T}_{in} \\ \tilde{S}_1 \\ \tilde{S}_2 \end{pmatrix}, w = \tilde{d}, u = \begin{pmatrix} \tilde{S}_1 \\ \tilde{S}_2 \end{pmatrix} \tag{5.11}$$

The scaled manipulated inputs $\tilde{S}_1$ and $\tilde{S}_2$ have been included in the regulated outputs to incorporate the input constraints on the splits. Zero steady-state offsets for T_{in} and T_{out} must be enforced. This involves adding two linear time domain constraints onto the resulting ℓ_1 LP. The objective of this problem is the same as that given in equation (3.32). The optimal objective ν_o should be less than 1 for both the performance specifications and input constraints to be met. The problem is a two-block problem, with more regulated outputs than control inputs, ie, $n_z > n_u$, which requires the use of the DA algorithm. This was solved using several different LP solvers (based on MINOS, CPLEX and OSL subroutines), which all gave the same results as would be expected. Overall we have found CPLEX to be the most robust of these routines.

The convergence of the solution is shown in Figure 5.5. This gives $\underline{\eta}_N$ for $N = 200$ as 2.0551 which means that the performance specifications can only be met, using an LTI controller, for about half the disturbance range.

To investigate how restrictive the input constraints on the control inputs, u, are on the achievable performance, the ℓ_1 analysis can be repeated excluding the control inputs from the regulated output z. Now the problem is one block $n_z = n_u$ and $n_y > n_w$ which means that the problem is good rank, ie, there are no rank interpolation conditions. This means that only zero interpolation conditions will be produced and the DA algorithm is not required. Constraints are still added to the LP to enforce zero steady-state offset. When the resulting LP was solved the solution was $\nu_0 = 1.8287$. The performance is improved only slightly by ignoring input constraints.

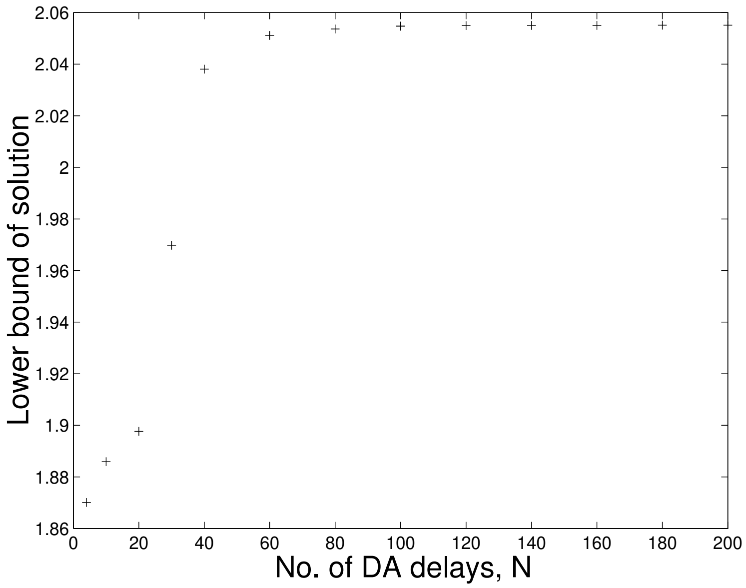

Figure 5.5: Convergence of DA algorithm for PFR for persistent disturbances

As stated previously, the disturbance can be rate limited through modifications of the upstream operation. Introducing a limit on the rate of change of the disturbance we found that the constraints could be satisfied (with no input constraints), if the disturbance varied only by 14.3% of its maximum variation per minute. With input constraints a variation of just 10% per minute could be tolerated.

The disturbance value was expected to change only infrequently, so a step description is arguably more appropriate. The problem in this case is still multi-block, since input constraints are still included, as are zero steady state conditions, and will require the DA algorithm.

The convergence of the DA for this problem is shown in Figure 5.6, giving $\underline{\eta}_N$ for $N = 200$ of 1.1428. Although this is a considerable reduction on the value for the ℓ_1 problem, it still does not quite satisfy the performance requirements and would need a further restriction of the disturbance set to do so. This could be done, as for the ℓ_1 analysis, by rate limiting the disturbance. This highlights the importance of the appropriate selection of the class of disturbances considered.

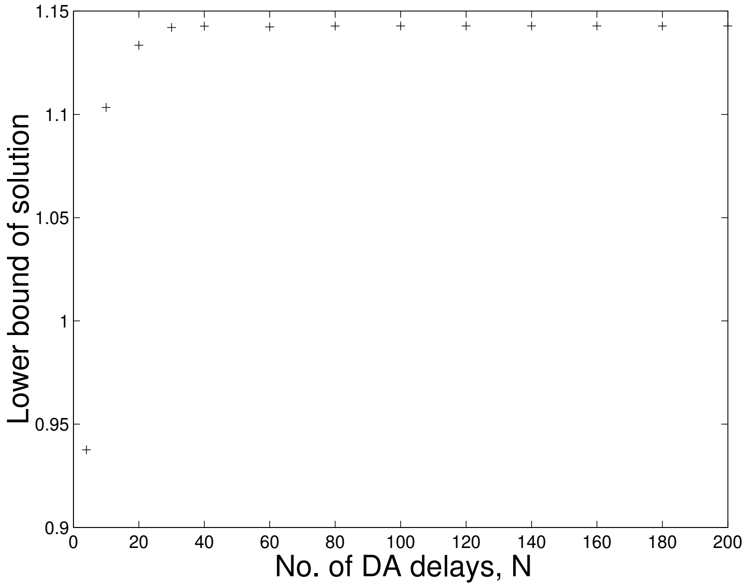

Figure 5.6: Convergence of DA algorithm for PFR for step disturbances

5.2.1 Conclusions

Overall this linear controllability analysis provides strong and unambiguous confirmation of the difficulty of meeting the performance requirements, particularly with rapidly varying disturbances.

The analysis carried out in Walsh and Malik (1995) indicated that a pessimistic bound on the achievable step disturbance rejection was violation of the required specification on T_{out} by a factor of 3. The actual nonlinear performance with a cascaded PI control scheme was noted to violate this limit by a factor of 1.5 for a smoothed step. The results from this linear optimisation based analysis are comparable. Both analyses highlight clearly that it is difficult to satisfy the specification.

5.3 Case Study 3: An Industrial Reactor System

This industrial reactor system consists of two adiabatic Plug Flow Reactors (PFRs) in series and is shown in Figure 5.7. A detailed study, involving analysis techniques not discussed here, is presented in Walsh *et al* (1997).

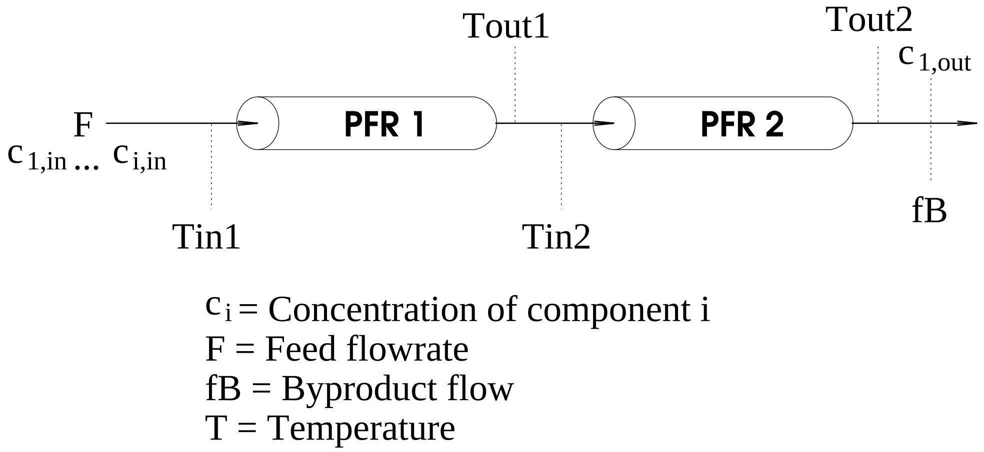

Figure 5.7: The industrial reactor system

The manipulable variables are $Tin1$ and $Tin2$, which are tightly regulated by secondary control loops. Measurement of the exit compositions on-line is expensive and will exhibit a significant delay of about 0.1 hours. Feed flow (F) and bed exit temperatures ($Tout1$, $Tout2$) can be measured reliably and rapidly.

The economic objective used was the minimisation of the production rate of a byproduct, fB, at the average (nominal) feed conditions. Therefore $J(u_0)$ in equation (4.2) is given by fB. This approximates to the expected value of fB over the disturbance set. For linear models and controllers this approximation is exact (see section 3.5.2). Process constraints apply to the reactor inlet and outlet temperatures and the product concentration of one component, $c_{1,out}$. The values are scaled, such that the allowable temperatures lie between 0 and 50 degrees. The natural log of the concentration $c_{1,out}$ ($lc_{1,out}$) has an upper limit of 2.2988 (the natural log is used to improve the accuracy of the linear models). The objective function, fB, is scaled so that a value of zero indicates the steady state economic optimum for the process with nominal feed properties.

The initial disturbance set to be considered was step variations in the feed flow and four inlet concentrations in the range of $\pm$ 10% of the nominal feed properties. This is obviously a crude estimate of the actual disturbances. However, it is sufficiently realistic to evaluate the effectiveness of controllability analysis techniques. Some of the analysis techniques used can only handle a restricted disturbance set, in which all changes are assumed to start from midway between the limits (the nominal feed properties in this case), so that steps from the lower to the upper limit are excluded. We therefore distinguish between a restricted step and a full step disturbance set in the results presented in the following.

A non-linear SPEEDUP model of the process was developed by ICI. The SPEEDUP model was converted to a gPROMS model to facilitate the use of dynamic optimisation.

A linearised model with 280 states was generated from the gPROMS model at the nominal steady-state optimum. For some of the techniques used it is important to use as simple a model as possible. A mixture of balanced truncation and residualisation was used to develop a 24 state approximation, given in Appendix D.1, of the original 280 state model. This model accurately matched the frequency response characteristics up to frequencies somewhat beyond the maximum anticipated bandwidth. The maximum bandwidth was estimated based on the frequencies at which the combined disturbance effect exceeded the combined control effect.

The initial linear model was found to give inaccurate predictions for $c_{1,out}$. This was due to this concentration exhibiting a markedly asymmetric response to positive and negative variations in inputs and disturbances. The natural log of this concentration $lc_{1,out}$ was predicted much more accurately by a linear approximation, so this variable was used to replace $c_{1,out}$ in the following analyses.

Initially qualitative analysis was carried out by studying the open-loop gain matrix of the linearised model:

$$\begin{bmatrix} fB \\ lc_{1,out} \\ Tout1 \\ Tout2 \end{bmatrix} = \begin{bmatrix} -10.221 & 3.596 & 1.233 & 2.145 \\ -5.715 & -3.1539 & 0.391 & -0.115 \\ 0.939 & 7.356 \times 10^{-5} & -7.173 \times 10^{-3} & 5.916 \times 10^{-2} \\ -1.22 \times 10^{-3} & 1.025 & -4.811 \times 10^{-3} & -7.742 \times 10^{-3} \end{bmatrix} \begin{bmatrix} Tin1 \\ Tin2 \\ F \\ c_{1,in} \end{bmatrix}$$

Increasing either $Tin1$ or $Tin2$ reduces $c_{1,out}$. Increasing $Tin1$ reduces fB, while increasing $Tin2$ increases fB. Qualitatively, the optimum strategy is, therefore, likely to be to adjust $Tin1$ to keep $Tout1$ on its limit and to keep $Tin2$ as low as its lower limit and the constraint on $c_{1,out}$ will permit.

Two disturbances show a dominant effect; feed flow (F) and one reactant concentration in the feed ($c_{1,in}$). This suggests that the other three disturbances could be neglected, giving a simpler analysis problem.

Before we apply the OLDE technique to the linear model it must be discretised. A sampling period of 0.0025 was chosen to avoid introducing significant additional bandwidth limits ($.0025 << \frac{\pi}{120}$, corresponding to a bandwidth limit of 120 rads/hour (Walsh *et al.*, 1997)). A horizon of 0.1 hours was used to include the full open loop effect of the disturbances. In this analysis, only the two dominant disturbances, feed flow and $c_{1,in}$, have been considered, so as to reduce the computational expense of the method. The full step disturbances are considered. The open-loop optimum for fB with the full step disturbance set is 1.207.

Applying this technique to the linear model with measured variables $Tout1$ and $Tout2$ and manipulated variables $Tin1$ and $Tin2$ the optimal value for fB was 0.100. Feed forward was added to the control structure by including the feed flow in the measured variables (feed flow has been identified as one of the dominant disturbances). The optimal value for fB was 0.076. Measuring only $Tout1$ and F gave an increase in fB to 0.108, indicating that $Tout2$ is providing useful information for control.

In the absence of measurement noise and uncertainty, clear potential for effective control performance with a linear multivariable feedback controller has been shown. Including feedforward information gives a moderate further improvement, suggesting that this possibility is worth considering. $Tin2$ is not manipulated by the controller in either case suggesting that this variable is redundant.

In the above analysis, the optimal controller is selected without any consideration of the effects of noise or uncertainty and without any requirement for a well-damped response. The controller generated is, therefore, not itself suitable for implementation and simply gives an estimate of the achievable performance with linear multivariable control.

Figure 5.8: The critical disturbance for the industrial reactor

To give some nonlinear validation, this problem is solved using the ONDE technique for one critical disturbance scenario. The critical disturbance scenario for $Tout1$ from the OLDE analysis was used to estimate the worst disturbance to be applied to the nonlinear dynamic model and is shown in Figure 5.8. In this figure 0 indicates the steady state and ± 1 indicates $\pm$ the maximum expected disturbance deviation. The disturbance starts from steady state, since for the ONDE method the restricted step disturbance set is used. This is necessary to introduce the nominal steady-state operating point, $(u_0,\ E(d))$, into the optimisation at t=0. There is no explicit controller to allow the closed loop steady-state behaviour to be defined more generally. The economic objective was minimised by optimising discrete values of $Tin1$ and $Tin2$, with the same sampling period and horizon as in the OLDE analysis. The optimal value of fB was 0.056 and the response of $Tout1$ is shown in Figure 5.9.

This result is better than the OLDE optimum of 0.076. Some sufficient reasons for this are given below. The nonlinear model is only subjected to one critical disturbance scenario, while in the OLDE analysis all disturbance scenarios are considered simultaneously. The disturbance scenario was selected from the restricted step disturbance set, rather than from the full step disturbance set used in the OLDE analysis. The optimal control is not required to be a function of an invariant control law. The constrained variables are not required to take a constant value at the end of the horizon.

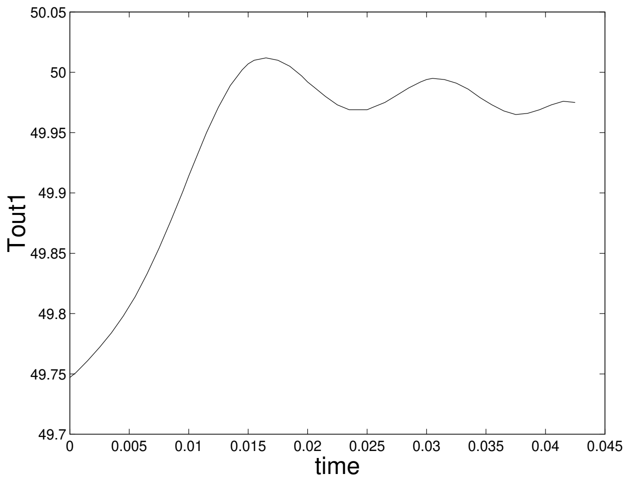

Figure 5.9: Initial response of $Tout1$ for the optimal control schedule

If the nonlinear optimal control problem gave an optimal value of fB greater than the OLDE value, then this would imply that, either the nonlinear global optimum had not been found, or the linear approximation used in the OLDE analysis was not sufficiently accurate.

The above nonlinear dynamic optimisation result provides a lower bound on the performance of any controller for which the steady-state operating cost for the average disturbance is representative of the overall operating economics.

The results show that advanced control has the potential to substantially improve performance compared to open-loop operation. The fact that the ONDE result does not substantially improve on the OLDE result suggests that LTI multivariable control can deliver most of the achievable benefits from advanced control.

5.3.1 Conclusions

The conclusions made, due to the results of the analyses carried out on this problem, depend on the significance of fB in real money. If one unit of fB was worth about £20,000 per year, then a further control study would be justified. The analysis so far would suggest that a LTI multivariable controller, adjusting just $Tin1$ using all available continuous measurements, should provide almost all the potential benefit from advanced control. However the disturbance characteristics should be reviewed carefully, as conclusions are very sensitive to the maximum rate of change of the disturbances. On the other hand, if one unit of fB was worth less than £2,000 per year, then this study should have stopped as soon as the open-loop dynamic economic analysis was completed.

The techniques used were successful in providing solid quantitative estimates of achievable performance. This allows an informed decision as to future action to be made. Also the linear analysis results provided excellent starting points for the non-linear analysis.

5.4 Case Study 4: An Evaporator System

The following case study is based on an evaporator model which is investigated in great detail in Newell and Lee (1989). The nonlinear model of this evaporator, shown in Figure 5.10, was programmed in gPROMS and has 20 variables, 12 equations and 3 states. The equations of the model are given in Appendix D.2. The level L2 is not self regulating, so for convenience a PI controller, designed in Newell and Lee (1989), is implemented as part of the model. This makes the optimisation easier with no loss of generality.

$$\Delta \mathrm{F2} = K_c(1 + \frac{1}{\tau_i s})\Delta \mathrm{L2}$$

where ΔL2=L2-L2$_{ss}$ and ΔF2=F2-F2$_{ss}$, for L2$_{ss}$=1 m and F2$_{ss}$=2 kg/min, and the constants are given as K_c=5.6 kg/min/m and τ_i=8.84 min.

There are two performance objectives for the controls in this case study. One is to maintain the deviations of the operating pressure, P2, and the product composition ,X2, within a certain range about their steady states, i.e.,

$$\begin{aligned} -\frac{range(\%)}{100\%}\mathrm{X2}_{ss} \leq \quad & \Delta \mathrm{X2} \quad \leq \frac{range(\%)}{100\%}\mathrm{X2}_{ss} \\ -\frac{range(\%)}{100\%}\mathrm{P2}_{ss} \leq \quad & \Delta \mathrm{P2} \quad \leq \frac{range(\%)}{100\%}\mathrm{P2}_{ss} \end{aligned} \tag{5.12}$$

These path constraints are implemented as shown in Appendix D.3. This range is not given in Newell and Lee (1989), but if $range(\%)$ were 0% then this implies perfect control, whilst if it were 100% this might imply poor control. Therefore the size of this range is a performance objective for this problem. Another performance objective is the optimal economics.

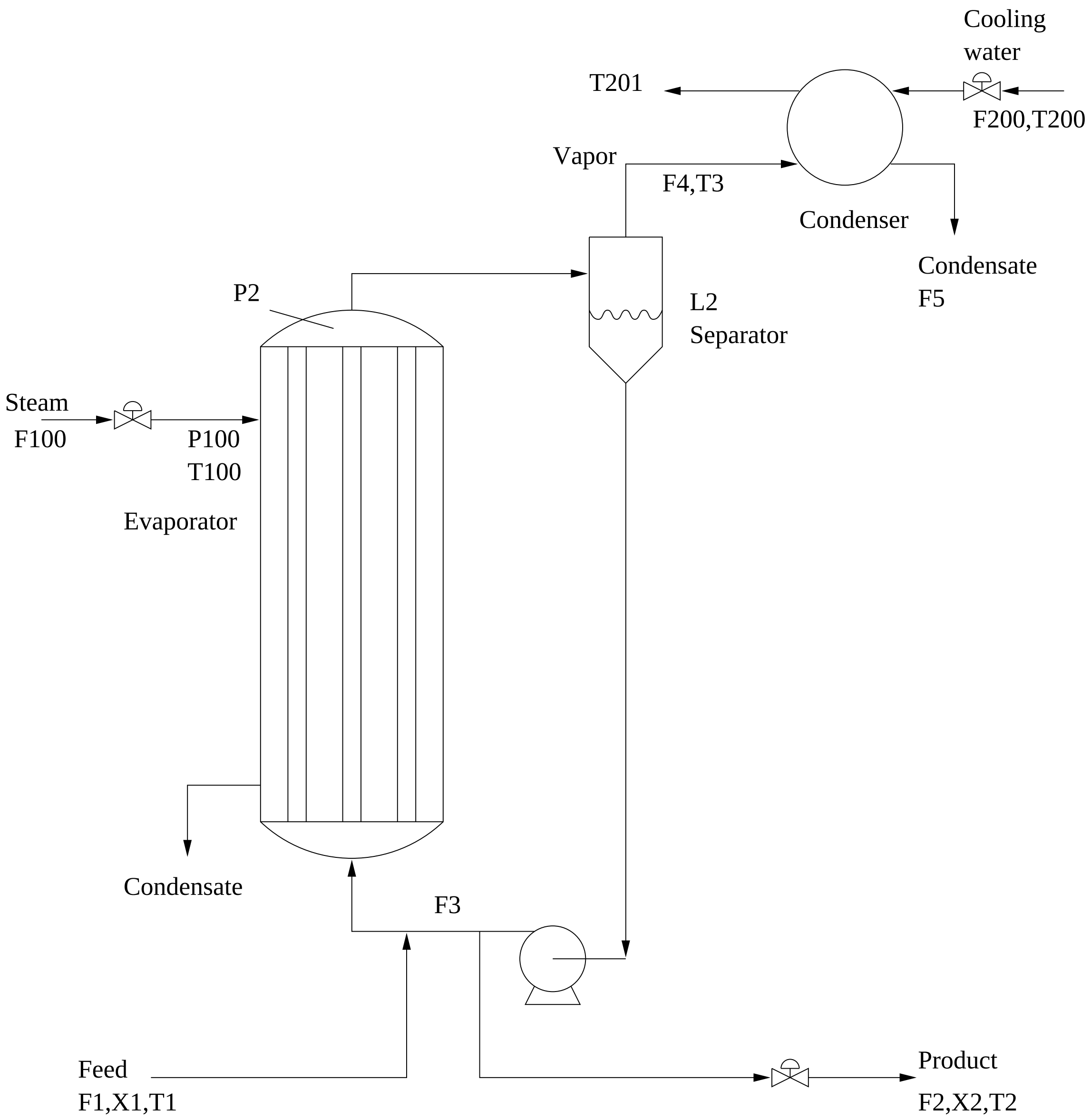

Figure 5.10: The evaporator system

The economics are given by the operating cost of the plant, which is assumed to be dominated by the steam and electricity consumption. Therefore Newell and Lee (1989) note that the economic performance is improved by minimising the recirculation and pressure. Insufficient information is provided to precisely weight these two variables, therefore as they are of similar magnitude we have chosen the objective function to be

$$\mathrm{F3}_{ss} + \mathrm{P2}_{ss} \tag{5.13}$$

subject to certain constraints. We want to maintain product quality at its operating point

$$\mathrm{X2}_{ss} = 25\% \tag{5.14}$$

which provides one operating point constraint. Also there are bounds on the steam pressure, P100, and the cooling water flowrate, F200, (the manipulated variables) which are used to control X2 and P2 (the measured variables) respectively.

$$\begin{aligned} \mathrm{P100} &\leq 400\mathrm{kPa} \\ \mathrm{F200} &\leq 400\mathrm{kg/min} \end{aligned} \tag{5.15}$$

The operating point of the circulating flowrate, $\mathrm{F3}_{ss}$, is selected to both minimise the objective and to satisfy the condition (5.14).

Using nonlinear steady state optimisation we get the optimal economic operating point to be

$$\begin{aligned} \mathrm{P100}^o_{ss} &= 400\mathrm{kPa} \\ \mathrm{F200}^o_{ss} &= 400\mathrm{kg/min} \\ \mathrm{F3}^o_{ss} &= 18.89\mathrm{kg/min} \end{aligned}$$

which give

$$\mathrm{P2}^o_{ss} = 18.89\mathrm{kPa}$$

and an objective, given by (5.13), of 59.38.

This means that P100 and F200 are sitting on their upper constraints. So, when a process disturbance occurs causing deviations in either X2 or P2, which require P100 or F200 to take positive control action to maintain (5.12), then these upper bounds will be violated. Therefore the operating point will have to be backed off its economic optimum sufficiently to allow the performance bounds in (5.12) to be satisfied, but still give the minimum possible operating costs.

Therefore the problem solved in this case study is the minimisation of the operating costs in (5.13), via the selection of P100, F200 and the operating point $F3_{ss}$, subject to the constraints in (5.12), (5.14) and (5.15) for the process disturbances:

$$\begin{aligned} F1_{ss} &\pm 10\% F1_{ss} \\ X1_{ss} &\pm 10\% X1_{ss} \end{aligned}$$

for different values of $range(\%)$, i.e., 10%, 20% and 30%. This will provide a picture of the trade-off between the two performance objectives.

The problem was first solved using the OLDE technique and then using the linear specialisation of the ONDE technique (see 4.8) to provide an upper and lower bound on the linear performance. These both require a linearised discrete model. The model was linearised about the optimal economic operating point and is given in Appendix D.4.

This linear continuous model was discretised using the Tustin method for a sampling period of 3 minutes. This sampling period T_s was based on the guidance, given in Newell and Lee (1989), that

$$T_s < 0.38 t_{min}$$

where t_{min} is the smallest time constant of interest. Using the linear continuous model described in Appendix D.4 t_{min} was identified as approximately 9.2 minutes (the identification was based on the Smith techniques described in Newell and Lee (1989)). Therefore $T_s < 3.5$ minutes. The problems in this case study are solved over a finite horizon of 300 minutes.

The results of the OLDE technique and the linear specialisation of the ONDE technique were found to correspond to within 0.01% of each other, therefore the linear results shall be discussed as one. It was found that for a value of $range(\%)$ of 10% both $P100_{ss}$ and $F200_{ss}$ were required to back off. However for values of 20% and 30% the open-loop deviations in P2 were well within the bounds, so no control via F200 was needed. This meant that only $P100_{ss}$ backed off. The results are shown in the following table.

performance bound	back off in $P100_{ss}$ ($\% \times P100^o_{ss}$)	back off in $F200_{ss}$ ($\% \times F200^o_{ss}$)	objective $F3_{ss}+P2_{ss}$
10%	-10.5%	-54.4%	67.709
20%	-9.9%	0%	61.808
30%	-3.1%	0%	60.146

It was found, that for a value of $range(\%)$ of 40%, no back off from the optimal economic operating point was required so the objective was 59.382. These two sets of linear results correspond so well, since investigation of the linearised model shows that the system has no NMP characteristics, this means that avoiding limitations due to the measured variables cannot help the idealised controller improve the performance. The main limitation on the achievable performance for this system is due to the input constraints on P100 and F200.

These results are plotted in Figure 5.11 where the objective is scaled as,

$$\left(\frac{\Delta F3_{ss}}{F3^o_{ss}} + \frac{\Delta P2_{ss}}{P2^o_{ss}}\right) \times 10 \tag{5.16}$$

where $\Delta F3_{ss}$ and $\Delta P2_{ss}$ are the back offs in the operating points $F3_{ss}$ and $P2_{ss}$ respectively.

Before carrying out the ONDE analysis with the nonlinear model, the linear specialisation of this technique was rerun with only 50 control intervals rather than 100. It was found that the results were within 0.01% of each other so the number of control intervals considered was reduced to 50 for the nonlinear analysis. Since nonlinear optimisation is computationally expensive, it is better to use the minimum required number of control intervals. Also the worst disturbances, for each constraint and for each range, were identified from the linear analysis by aligning the peaks of the resulting closed-loop step responses to give the worst deviations. It was found that for all the ranges, i.e., 10%, 20% and 30%, and all the constraints the worst disturbance was given by the disturbance shown in Figure 5.12. There is no delay between the occurrence of F1 and X1, so that means that the acausal nature of the idealised controller can be avoided.

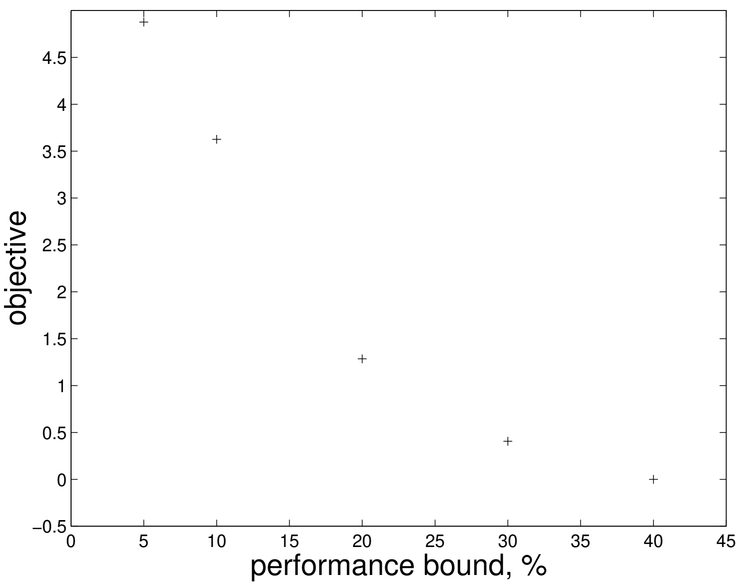

Figure 5.11: Objective vs. performance bounds for the linear model

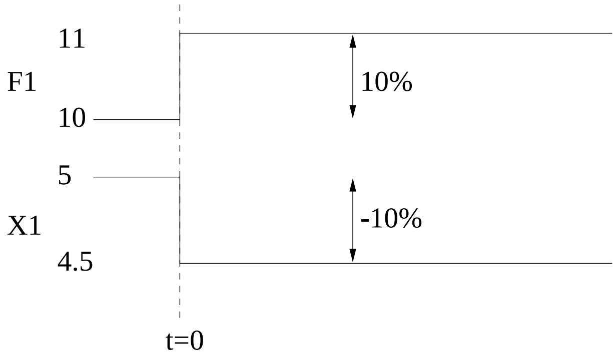

Figure 5.12: The Worst Disturbance for the evaporator system

The gPROMS model was run for this disturbance and the responses of P2 and X2 are shown in Figure 5.13. It can be see that open-loop P2 increases by about 8.3% and X2 decreases by about 27.5%. Clearly a value of $range(\%)$ of 30% will require no control and therefore no back off.

In a similar manner to the linear analysis for a value of $range(\%)$ of 10% both $P100_{ss}$ and $F200_{ss}$ were required to back off, while for 20% the control F200 was unused so only $P100_{ss}$ was backed off. The results are shown in the following table.

performance bound	back off in $P100_{ss}$ ($\% \times P100^o_{ss}$)	back off in $F200_{ss}$ ($\% \times F200^o_{ss}$)	objective $F3_{ss}+P2_{ss}$
10%	-10.3%	-33.1%	69.228
20%	-7.8%	0%	61.418
30%	0%	0%	59.382

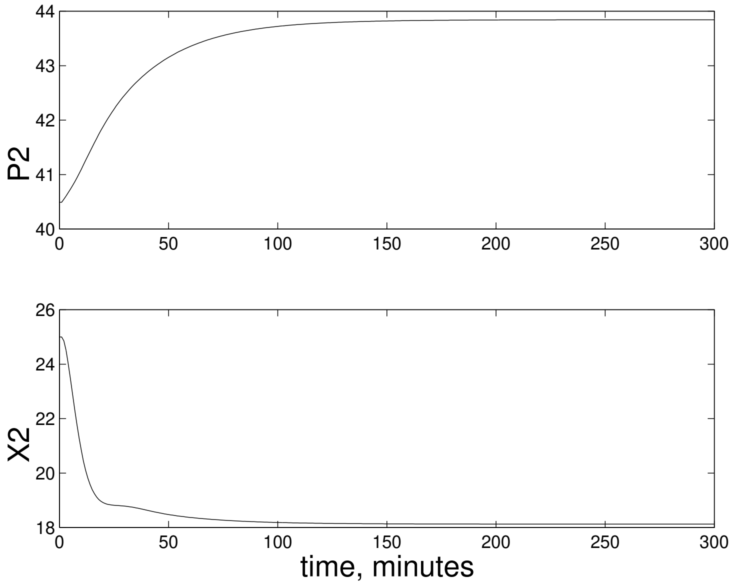

Figure 5.13: The responses of X2 and P2 to the worst disturbance

These results are scaled in the same manner as that described in (5.16) for the linear case and are shown in Figure 5.14.

To run the analysis ($range(\%)$=20%) for more than one disturbance the step in X1 was delayed by one control interval and added to the disturbance set. Unsurprisingly the performance was found to be the same since the best achievable objective for this delayed disturbance alone is marginally better, as shown in the following table.

delay in X1 (minutes)	objective
0	61.41812
6	61.41788
12	61.41681

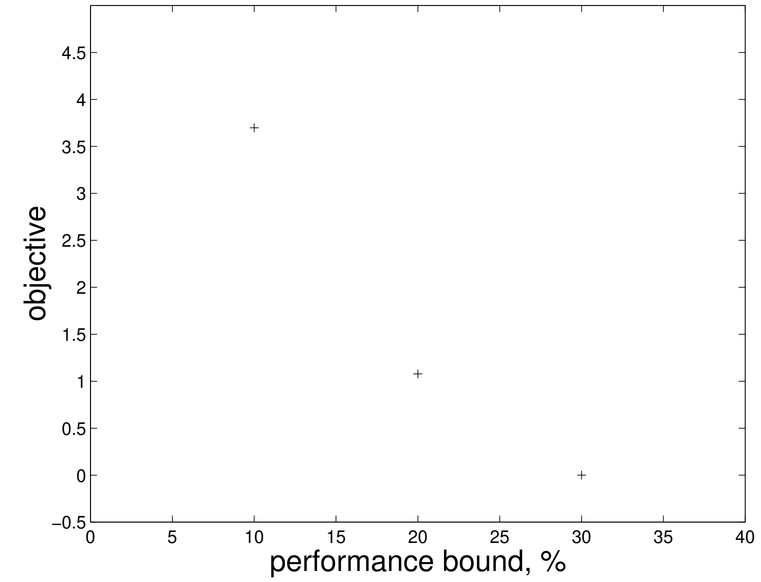

Figure 5.14: Objective vs. performance bounds for the nonlinear model

The technique for limiting the acausal nature of the idealised controller was applied to this delayed disturbance. However, a candidate disturbance for $\tilde{w}^o$, which gives a constraint violation when $\tilde{u}_b = u_b^o$, as described in 4.5.1, could not be found. Therefore this is an example of when the technique for limiting acausal behaviour cannot find an appropriate disturbance.

5.4.1 Conclusions

The linear and nonlinear objectives correspond well for a value of $range(\%)$ of 20%, with the nonlinear results giving a slightly better performance for lower values of $range(\%)$ and a slightly worse performance for higher values of $range(\%)$. All the analysis techniques show that the achievable economic performance improves as the performance bounds on the deviations in X2 and P2 are reduced. In fact, if we allow these bounds to be greater than 30% of their steady states, then the nonlinear analysis shows that the plant can operate at the optimal economic operating point.

In Newell and Lee (1989) both P100 and F200 are backed off to 350 to allow for control action, giving a $F3_{ss}$ of 22.8 kg/min and $P2_{ss}$ of 42 kPa. For the disturbance description used here, if the bounds on the deviations in X2 and P2 were 20%, then this back off would be excessive and the nonlinear analysis in this section suggests that an economically better operating point of F3$_{ss}$=20.9 and P2$_{ss}$=40.5 might be achieved. However, if the performance bounds were tighter, i.e., 10%, then the analysis suggests there might be constraint violations for these operating points of 350 and that, in fact, there might have to be a further back off to maintain feasibility.

The specifications of the process disturbances and the required performance bounds on the deviations in X2 and P2 would have to be clarified before this analysis could be used to draw any specific conclusions about the achievable performance.

5.5 Conclusions

The software has been validated against published results and applied to a couple of industrial process control problems and a well published process control problem. The results confirm the utility of the method, but highlight some limitations and some general questions for further consideration.

Perhaps the major limitation of all the technique presented here, is their computational expense. The linear method was solved using the interpolation conditions, which required a lot of computation simply to set up the LP. The series of LPs in the first example solved to give the DA convergence quite quickly, whilst the second example took a couple of hours to give the full convergence curve shown. The number of variables (n_z, n_w, n_u, n_y), the length of the finite horizon and the number of different values of the DA delay, N, affect how long this technique takes. The third example has a large number of variables and states, so that, even for a fairly short finite horizon, this example took several hours to solve using the DA algorithm. The solution time could be reduced by solving for only a few values of N and for a reduced finite horizon.

On the other hand, the linear specialisation of the nonlinear controllability technique solved quickly and, on the whole, gave a solution within several minutes.

The nonlinear method is computationally expensive due to the need for nonlinear dynamic optimisation. The solution time varied widely depending on the size of the nonlinear model, i.e., the number of states, variables and equations, and the number of control intervals used. The third example had in the order of a few hundred states and a few thousand variables and equations, making the nonlinear model large and unwieldy. This was solved for 40 control intervals (using the same sampling as the linearised problem) and took several days to solve. However the last example had 20 variables, 12 equations and 3 states and only took an hour or so to solve for 50 control intervals. (All the solution times discussed here are based on a SPARC 10 workstation).

Other issues include, the discretisation of the model for the linear techniques and the resulting number of control intervals used for the nonlinear method, and the use of different step disturbance descriptions, i.e., full or restricted.

The sampling level of the discretisation should be selected to be high enough that the achievable performance as closely resembles that of the continuous time system as possible. If the system is bandwidth limited, then selecting the sampling frequency to be greater than twice the bandwidth limit frequency means that the continuous system is completely represented by the discrete system (the sampling theorem see (Oppenheim *et al.*, 1983)). If the system is not bandwidth limited, then some maximum frequency of interest must be estimated and the discrete system will be an approximation of the continuous system. This sampling frequency will also limit the achievable controller, since it will not be able to have any higher frequency dynamics, than those stipulated by the sampling frequency. As this sampling frequency is increased, then the closer the set of available discrete controllers represents the set of continuous controllers. However, as the sampling frequency increases, so does the size of the optimisation problem to be solved, for the same finite horizon. If the same sampling period is used as the control interval for the nonlinear problem, then it is particularly important to keep this as small as possible.

The linear method uses a full step description (stepping from one bound to the other), whilst the nonlinear method and its linear specialisation use a restricted step description (stepping from its steady state to one or the other bound). This means, that the performance for the full step may be worse than that for the restricted step, depending on the specific system. For a system which only displays first order behaviour this will make no difference. Since the techniques using the restricted step description are already lower bounds on the achievable performance, it is not too serious a matter that this disturbance description may make them yet more optimistic. These idealised controller techniques have to initiate from their expected operating point for the performance to be estimated at this point, therefore not much can be done to alter these methods. However the LTI controller technique can easily be adapted to restricted step disturbances, so that the correspondence between this result and the idealised controller results could be checked if the achievable performances differed greatly. This would allow the effect of the use of a full or restricted step disturbance description to be assessed for a particular problem.

Chapter 6

Conclusion

The aim was to develop controllability analysis techniques, for both linear and nonlinear models, which give unambiguous measures of achievable control performance prior to the design of the controller. Desirable objectives for any such method are that it should:

- capture typical performance requirements directly;
- find the best performance with a class of controllers, which is as broad and realistic as possible;
- take account of as many of the fundamental limitations on controllability as possible;
- be computationally tractable.

A method for the linear controllability analysis has been developed that meets many of the above objectives. The complementary nonlinear controllability analysis technique falls short on some of these characteristics, but, as far as possible, attempts to provide a similar controllability measure for the nonlinear problem, as the linear technique provides for the linear problem.

Both methods can be formulated to capture many typical primary performance requirements in a very flexible manner without need for approximation. Specifically the OLDE/ONDE problems capture the best achievable economic performance, by evaluating the minimum back off from the economically optimal operating point required to ensure that none of the process constraints are violated for any of the process disturbances. The closer this new feasible operating point is to the optimal operating point, the better the economic performance. If we don't have information about the economics of the process, then the methods might be used to evaluate the feasibility of the problem directly or to assess the "disturbance fraction" which the process can tolerate and for which it can still maintain feasibility. Little attention has been paid to classical control criteria, such as damping and integral squared error. This reflects their secondary importance, the difficulty of weighing such issues objectively against economic objectives, and, for the linear method, the difficulty of including them within an LP framework. The controllers obtained from the optimisations may, therefore, be lightly damped or otherwise "poor" controllers. We prefer to let a control engineer make the trade-off between "nice" responses and economics/ feasibility, with the knowledge of the best that can be achieved ignoring such issues.

Both techniques are based on optimal control problems and, therefore, the more general and realistic the set of controllers provided to the optimisation, the more realistic is the measure of best achievable performance. If the set is a realistic, but restricted set, i.e., the set of PI controllers, then the result can be achieved using an implementable controller, however the truly optimal controller may not be included in this set. Therefore the result, in this case, is a pessimistic bound on the best achievable performance. On the other hand, if the set of possible controllers is an idealised set, i.e., perfect control, then the optimal controller is not realisable and any implementable controller would not be able to achieve this level of performance. In this case the result is an optimistic bound on the best achievable performance. It must be an objective of any controllability technique to try to provide a result which is as tight a bound on the best achievable performance as possible. Therefore the selection of the controller set is of great importance to any controllability analysis technique based on optimal control. For the linear controllability technique the set of controllers is the set of all stabilising LTI controllers, which is both a broad and meaningful one, but it is not complete. The most common, conventional and advanced controllers in the process industry, PI with output limiting and constrained model predictive control, lie outside the set of LTI controllers as their characteristics change when constraints become active. Therefore the result produced by the linear controllability analysis is a pessimistic bound, since the optimal controller may not be included in this set. Both the nonlinear controllability method and its linear specialisation optimise the performance by solving an optimal idealised control problem. The controller is idealised, since no limitations associated with measurement availability and characteristics are captured and the controller is acausal. The last point is due to the optimiser having full information on the disturbance before it selects the optimal control schedule. However a technique has been implemented to limit the acausal nature of these optimal controllers in an attempt to make the optimistic bound, that this result produces, tighter. If the pessimistic bound, produced by the linear controllability method, and the optimistic bound, produced by the linear specialisation of the nonlinear controllability method, are close, then this suggests that these bounds are tight. The nonlinear controllability analysis technique can then be applied to the nonlinear model and the results compared, to ensure a good correlation between the two models and to validate the linear results.

Limitations on the achievable performance due to input constraints and noise can be included in both the linear and nonlinear methods with no problem. The effect of time delays and unstable zero dynamics in the system are incorporated directly in the linear technique through the feasibility constraints. However, although the effect of these occurring between the control inputs and the regulated outputs are automatically captured in the nonlinear technique, any limitations due to these occurring between the disturbances and the measured variables are excluded due to the nature of the idealised controller used. In fact, this controller will actually have information of the disturbance before it has occurred, therefore the nonlinear method has been adapted to at least attempt to ensure a degree of causality. The one fundamental limitation on controllability, which is excluded from both techniques completely, is uncertainty. As discussed in Dahleh and Diaz-Bobillo (1995) it is straightforward to add unstructured robust stability conditions to the linear controllability problem at the expense of increasing problem size. Such descriptions are however notoriously conservative. Dealing with robust performance or structured uncertainty, in a non-conservative manner, for the linear method requires further research. One possibility would be to use the linear method as a guide to developing a controller and subsequently address robustness directly on the nonlinear model. There already exist several nonlinear optimal control problems that assess performance for uncertain models, as discussed in section 2.3, therefore further work might be undertaken to augment the nonlinear controllability method to incorporate uncertainty. However this would lead to a more complex and therefore more computationally expensive problem.

The computational tractability has been verified to some extent by successful performance on industrial problems. Nonetheless there are limitations on the size of problem which can be analysed at present. For the linear method, a problem with $n_z = n_u = n_y = n_w = n = 5$ and a finite horizon of 100 would give several thousand constraints and variables, close on a million non-zero elements and take several hours to solve on a SPARC 10 workstation. Simple linear programming algorithms will fail on a problem of this size. In the short-term careful formulation and the use of high quality LP software is necessary to allow realistic problems to be tackled. The nonlinear controllability method is computationally more expensive than the linear method. A problem with around 50 control intervals, a few thousand variables and equations and a few hundred states took several days to solve on a SPARC 10 workstation, whilst a problem with less than 50 variables and equations and less than 10 states would take around a couple of hours. The number of control intervals used must be selected with the size of the problem in mind, since the solution time will increase with this number. Also the solution time is greatly improved by the selection of a good initial guess of the optimal point. As both nonlinear and linear optimisation techniques improve, and computer power is increased, then larger and more complex problems will become solvable.

These two techniques are most useful when used in conjunction. The linear controllability technique provides useful estimates for the worst disturbance for the nonlinear controllability analysis. The nonlinear technique provides validation for the linear controllability analysis results.

The result provided by the linear controllability analysis is an upper bound on the best achievable control performance for the linearised nominal process. The nonlinear controllability analysis provides a lower bound on the best achievable control performance for the nominal process. The linear specialisation of the nonlinear controllability technique can be used to provide a lower bound on the linear performance. If the nominal process:

- passes all these tests, then it implies that the nominal process is controllable and that fixed linear feedback may be adequate to achieve the required performance.
- fails the linear controllability analysis, but passes the optimal idealised control based techniques, then this implies that the process is potentially controllable, but probably will require a nonlinear controller or a modified measurement set.
- fails both the linear controllability analysis and the linear specialisation of the nonlinear controllability analysis, but passes the nonlinear test, then this implies that the nonlinearities assist the control and therefore all control design should be based on the nonlinear model.
- fails all the tests, then this implies that either the nominal process is not controllable or that the NLP has failed to find the global optimum.

Other methods could be used to complement these techniques. For example, nonlinear optimisation of simple implementable controllers might be used to give an upper bound on the nonlinear performance, or controllability indicators could be used to highlight characteristics contributing to poor performance and to aid a search for better process and control structure options.

The contributions of this thesis are the formulation and implementation of optimal control problems to give controllability measures and their application to realistic industrial problems. The linear controllability analysis has required the development of formulations which allow the problem to incorporate a range of realistic disturbance descriptions (section 3.4) and typical performance requirements (section 3.5). The ℓ_1 performance problem is an existing optimal control method (Dahleh and Diaz-Bobillo, 1995) and the techniques used for solving it are used as a basis for the linear controllability analysis developed here. A specific application of the linear controllability analysis technique has been formulated which assesses the economic performance of a linear dynamic system based on back offs, the OLDE technique (section 3.5.2). Working in parallel, Young *et al* (1996) have presented a similar approach which differs significantly in the details. For the nonlinear controllability analysis, formulations have been developed to allow the evaluation of the performance over a set of multiple step disturbances (section 4.4) and the incorporation of typical performance requirements (section 4.6). The ONDE technique has been formulated (section 4.6.1) to give a measure of the economic performance in a similar manner to the OLDE technique, but for nonlinear dynamic systems. A technique is proposed to tighten the lower bound provided by this nonlinear method by limiting the acausal nature of the idealised controller (section 4.5.1). The nonlinear controllability technique has been reformulated for the special case when it is applied to a linear model (section 4.8). In this case it can be solved extremely efficiently as an LP. A simple linear example is presented demonstrating this linear specialisation and the technique for limiting acausal behaviour. All these controllability analysis techniques have been implemented and applied to industrial problems (see Chapter 5) of realistic size.

Some suggested areas for further work, relating to this thesis, are as follows:

- the development of an upper bound on the nonlinear performance, e.g., using an optimal realisable (PI or LTI) control problem.
- further tightening of the lower bound provided by the optimal idealised control problem.
- the incorporation of uncertainty into the controllability techniques.

- further application of the technique for limiting acausal behaviour to nonlinear problems.
- further investigation into the effect of different discretisation techniques and sampling periods on the measure of performance.

Finally there is always a need for computational improvements with such techniques, to shorten the solution time and to diminish the computational expense.

Appendix A

Finite Interpolation Conditions

The interpolation conditions consist of rank interpolation conditions and zero interpolation conditions. The former produce an infinite number of constraints and variables, so an algorithm for avoiding the use of these constraints is described in A.1.The latter produce an infinite number of variables and a technique for truncating the variable vector to a finite length is described in A.2.

A.1 The DA Algorithm

As mentioned earlier one-block problems have no rank interpolation conditions, therefore this fact can be exploited to avoid solving the infinite constraint and variable linear programs produced by multiblock problems. There are several methods of truncating the original problem to give a one-block problem with only zero interpolation conditions. In this work the Delay Augmentation (DA) has been used. This embeds the problem in a one-block problem through augmenting the operators $U$and V with delays. The optimal solution of this augmented one-block problem gives a lower bound on the optimal solution of the original problem. The optimal Q_N^o for the augmented problem allows us to form an upper bound. The delay is increased until these two bounds have converged sufficiently.

If we partition the original problem as follows,

$$\begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} - \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} Q \begin{pmatrix} V_1 & V_2 \end{pmatrix} \tag{A.1}$$

where $U_1 \in \ell_1^{n_u \times n_u}$ and $V_1 \in \ell_1^{n_y \times n_y}$, then the DA problem would be given by

$$\begin{pmatrix} \Phi_{11,N} & \Phi_{12,N} \\ \Phi_{21,N} & \Phi_{22,N} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} - \begin{pmatrix} U_1 & 0 \\ U_2 & S_N \end{pmatrix} \underbrace{\begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}}_{Q_N} \begin{pmatrix} V_1 & V_2 \\ 0 & S_N \end{pmatrix} \tag{A.2}$$

where S_N are Nth order forward shifts. The statement given for the ℓ_1 case in Dahleh and Diaz-Bobillo (1995) on using the DA algorithm to bound ν_o can be made more generally.

For the performance problem

$$\min_{K,u_0} J(K,u_0) \;\; s.t. \;\; c(K,u_0,w) \leq 0 \; \forall \; w \in W \tag{A.3}$$

let $\tilde{J}(Q,u_0) = J(K,u_0)$ and $\tilde{c}(Q,u_0,w) = c(K,u_0,w)$ so that A.3 can be stated as

$$\min_{Q \in \ell_1^{n_u \times n_y}, u_0} \tilde{J}(Q,u_0) \;\; s.t. \;\; \tilde{c}(Q,u_0,w) \leq 0 \; \forall \; w \in W \tag{A.4}$$

In this case the optimal solution for the delay augmented problem is given by

$$\underline{\eta}_N = \min_{Q_N \in \ell_1^{n_z \times n_w}, u_0} \tilde{J}(Q_N,u_0) \;\; s.t. \;\; \tilde{c}(Q_N,u_0,w) \leq 0 \; \forall \; w \in W \tag{A.5}$$

whilst the optimal solution for the original problem is given by (A.4) and can rewritten as

$$\nu_o = \min_{\substack{Q_{11} \in \ell_1^{n_u \times n_y}, \\ Q_{12}=Q_{21}=Q_{22}=0, u_0}} \tilde{J}(Q_N,u_0) \;\; s.t. \;\; \tilde{c}(Q_, u_0,w) \leq 0 \; \forall \; w \in W \tag{A.6}$$

therefore the extra degree of freedom provided by Q_N allows the construction of super optimal solutions, ie,

$$\underline{\eta}_N \leq \nu_o. \tag{A.7}$$

This gives a lower bound $\underline{\eta}_N$ to the optimal solution of the original problem ν_o.

When N has been increased to a point at which a Q_{11}^o is found which satisfies

$$\tilde{c}(Q_{11}^o, \bar{u}_0, w) \leq \varepsilon \; \forall \; w \in W \tag{A.8}$$

where Q_{11}^o is the value of Q_{11} that gives the optimal value $\underline{\eta}_N$ of the augmented problem (A.5), $\bar{u}_0$ is chosen to satisfy (A.8) and ε is positive and small. An upper bound is given by

$$\begin{aligned} \nu_o &= \min_{Q\in\ell_1^{nu\times ny},u_0} \tilde{J}(Q,u_0) \;\; s.t. \;\; \tilde{c}(Q,u_0,w)\leq 0\;\forall\; w\in W && \text{(A.9)} \\ &\leq \tilde{J}(Q_{11}^o,\bar{u}_0) = \overline{\eta}_N && \text{(A.10)} \end{aligned}$$

In which case we have both upper and lower bounds on ν_o.

$$\underline{\eta}_N \leq \nu_o \leq \overline{\eta}_N. \qquad \text{(A.11)}$$

A.2 Truncation of the Infinite Variable Vector

The zero interpolation conditions produce an infinite number of constraints due to length of ϕ. We can rearrange the columns of the matrix A_{zero} to give a set of matrices M_{ij} for $i = 1,..,n_z$ and $j = 1,..,n_w$ where

$$(M_{ij})^k = (A_{zero})^{n_z n_w k + N_w(i-1)+j}$$

,ie , the column of A_{zero} that multiplies $\phi_{ij}(k)$. Therefore $A_{zero}\phi = b_{zero}$ can be rewritten as

$$\sum_{i=1}^{n_z}\sum_{j=1}^{n_w} M_{ij}\phi_{ij} = b_{zero}.$$

Clearly if we truncate the discrete impulse response of ϕ_{ij} to a length of N_{ij} so that

$$\sum_{i=1}^{n_z}\sum_{j=1}^{n_w}\sum_{k=0}^{N_{ij}} (M_{ij})^k \phi_{ij}(k) = b_{zero}$$

then

$$\sum_{i=1}^{n_z}\sum_{j=1}^{n_w}\sum_{k=N_{ij}+1}^{\infty} (M_{ij})^k \phi_{ij}(k) = 0.$$

This means that a valid choice of ϕ_{ij} is

$$\phi_{ij} = \begin{pmatrix} \tilde{\phi}_{ij} \\ \underline{0} \end{pmatrix}$$

where $\tilde{\phi}_{ij}$ is the truncated ϕ_{ij}.

There is a trick that can be used for the ℓ_1-optimisation problem specifically that allows the truncation of ϕ_{ij} without causing any further restriction on the set of ϕ from which the optimisation can choose. It makes use of a variable change that is made to enforce the absolute norm involved in the objective function, $\Phi = \Phi^+ - \Phi^-$ for $\Phi^+, \Phi^- >= 0$. The setting up of the ℓ_1-optimisation problem is described in more detail in 3.5.1.

The primal problem is given by

$$\nu_o = \inf_x \langle x, c \rangle \tag{A.12}$$

subject to

$$Ax = b,$$
$$x \geq 0$$
$$x \in \mathcal{R} \times \mathcal{R}^{n_z} \times \ell_1 \times \ell_1$$

where $x := (\nu\ \xi^T\ (\phi^+)^T\ (\phi^-)^T)^T$, $A_{\ell_1}(\phi^+ + \phi^-) + \xi = \mathbf{1}\nu$, ξ = slack variables, $c := (1\ 0\ 0....)^T$,

$$A := \begin{pmatrix} \mathbf{-1} & I & A_{\ell_1} & A_{\ell_1} \\ \mathbf{0} & \mathbf{0} & A_{zero} & -A_{zero} \end{pmatrix}, b := \begin{pmatrix} \mathbf{0} \\ b_{zero} \end{pmatrix}. \tag{A.13}$$

The problem has an infinite number of variables $x := (\nu\ \xi\ \phi^+\ \phi^-)^T$ due to ϕ^+ and ϕ^- just as discussed previously. However by considering the dual problem it can be shown that the underlying problem is finite dimensional.

The dual problem has a finite number of variables, but an infinite number of constraints.

$$\nu_o = \max_\gamma \langle b, \gamma \rangle$$

subject to

$$A^T\gamma \leq c, \tag{A.14}$$
$$\gamma \in \ell_\infty.$$

If we put $\gamma =: (-\gamma_0 \; \gamma_1)^T$ then the problem becomes

$$\nu_o = \max_{\gamma_0, \gamma_1} \langle b_{zero}, \gamma_1 \rangle$$

subject to

$$\gamma_0 \geq 0, \quad \sum_{i=1}^{n_z} \gamma_0(i) \leq 1, \tag{A.15}$$

$$-A_{\ell_1}^T \gamma_0 \leq A_{zero}^T \gamma_1 \leq A_{\ell_1}^T \gamma_0$$

$$\gamma_0 \in \mathcal{R}^{n_z}, \gamma_1 \in \mathcal{R}^{c_z}.$$

The last line of constraints can be restated using the M_{ij}'s presented earlier.

$$-\gamma_0(i) \leq (M_{ij}^T \gamma_1)(k) \leq \gamma_0(i) \quad for \quad \begin{cases} i = 1, ..., n_z \\ j = 1, ..., n_w \\ k = 0, 1, ... \end{cases} \tag{A.16}$$

It can be shown that for some N_{ij},

$$\|(I - P_{N_{ij}}) M_{ij}^T x\|_\infty < \|P_{N_{ij}} M_{ij}^T x\|_\infty$$

(for proof see (Dahleh and Diaz-Bobillo, 1995) p.272). Therefore, for $x = \gamma_1$, if $|(M_{ij}^T \gamma_1)(k)| \leq \gamma_0(i)$ for $k = 0, 1, ...N_{ij}$ then

$$\|(I - P_{N_{ij}}) M_{ij}^T \gamma_1\|_\infty < \|P_{N_{ij}} M_{ij}^T \gamma_1\|_\infty \leq \gamma_0(i). \tag{A.17}$$

This means that so long as the constraint is satisfied for $k = 0, ...N_{ij}$ then the constraint is inactive (ie, already satisfied) for $k > N_{ij}$. Therefore using an algorithm to find $N_{ij} \;\; \forall \, 1 \leq i \leq n_z, \; 1 \leq j \leq n_w$ these constraints become

$$-\gamma_0(i) \leq (M_{ij}^T \gamma_1)(k) \leq \gamma_0(i) \quad for \quad \begin{cases} i = 1, ..., n_z \\ j = 1, ..., n_w \\ k = 0, 1, ...N_{ij} < \infty. \end{cases} \tag{A.18}$$

So the dual linear program now has a finite number of variables (ie, $n_z + c_z$) and a finite number of constraints (ie, $n_z + 1 + \tilde{c}_z$).

Note: P_N=truncation operator, ie, $P_N x(k) = (x(0), x(1), ..., x(N), 0, 0....)$ for $x(k) = (x(0), x(1), ...)$, and

$$\tilde{c}_z = \sum_{i=1}^{n_z} \sum_{j=1}^{n_w} N_{ij} + 1. \tag{A.19}$$

The finite primal linear program can now be given by,

$$\nu_o = \min_{\tilde{x}} \langle \tilde{x}, c \rangle \tag{A.20}$$

subject to

$$\tilde{A}\tilde{x} = \tilde{b}, \tag{A.21}$$

$$\tilde{x} \geq 0$$

where $\tilde{x} = (\nu\ \xi^T\ (\tilde{\phi}^+)^T; (\tilde{\phi}^-)^T)^T$, $c := (1\ 0\ 0....)^T$, for $\tilde{\phi}_{ij}$ =finite vector consisting of the first $N_{ij}+1$ elements of ϕ_{ij},

$\tilde{\phi} = ((\tilde{\phi}_{11})^T....(\tilde{\phi}_{1n_w})^T\ (\tilde{\phi}_{21})^T....(\tilde{\phi}_{2n_w})^T....(\tilde{\phi}_{n_z 1})^T....(\tilde{\phi}_{n_z n_w})^T)^T$,

$$\tilde{A} := \begin{pmatrix} \mathbf{-1} & I & I_{\ell_1} & I_{\ell_1} \\ \mathbf{0} & \mathbf{0} & M_{zero} & -M_{zero} \end{pmatrix}, \tilde{b} := \begin{pmatrix} \mathbf{0} \\ b_{zero} \end{pmatrix}. \tag{A.22}$$

define the matrix I_{ℓ_1} such that,

$$(I_{\ell_1}\tilde{\phi})_i = \sum_{j=1}^{n_w} \sum_{k=0}^{N_{ij}} \tilde{\phi}_{ij}(k) \tag{A.23}$$

and M_{zero} such that

$$M_{zero}\tilde{\phi} = \sum_{i=1}^{n_z} \sum_{j=1}^{n_w} \tilde{M}_{ij}\tilde{\phi}_{ij} \tag{A.24}$$

for $\tilde{M}_{ij}$ = finite matrix consisting of the first $N_{ij}+1$ columns of M_{ij}. This can now be solved as a standard linear program to give the optimal value of $\|\Phi_N\|_1$ as N is increased.

Appendix B

Linear Implementation

Layouts for the linear software implemented for the interpolation technique and the Q-parametrisation technique are shown in Figures 3.3 and 3.4 respectively. Details for the computation of the Youla parametrisation and the null chains, as well as a technique for reconditioning part of the LP, are given in the following appendices.

B.1 The Youla parametrisation

The Youla parametrisation of the controller allows the lower LFT, Φ, for a stabilising controller, to be stated as an affine function of the stable parameter Q.

$$\Phi = H - UQV \tag{B.1}$$

where

$$\begin{aligned} H &= P_{11} + P_{12} Y \tilde{M} P_{21} \\ U &= P_{12} M \\ V &= \tilde{M} P_{21}. \end{aligned} \tag{B.2}$$

Rather than have to directly calculate the right and left coprime factorisations $N, M, \tilde{N}, \tilde{M}$ and the related $X, Y, \tilde{X}, \tilde{Y}$ we can use the following theory (see (Green and Limebeer, 1995) p.458) to develop H, U and V.

Theorem B.1.1 *Suppose $G = D + C(sI - A)^1 B$ is a stabilisable and detectable realisation of a transfer function matrix G. Let F be a state-feedback gain matrix such that $A - BF$ is asymptotically stable and let L be an observer gain matrix such that $A - LC$ is asymptotically stable. Define*

$$\begin{bmatrix} M & Y \\ N & X \end{bmatrix} = \left[\begin{array}{c|cc} A - BF & B & L \\ \hline -F & I & 0 \\ C - DF & D & I \end{array}\right] \tag{B.3}$$

$$\begin{bmatrix} \tilde{X} & -\tilde{Y} \\ -\tilde{N} & \tilde{M} \end{bmatrix} = \left[\begin{array}{c|cc} A - LC & B - LD & L \\ \hline -F & I & 0 \\ -C & -D & I \end{array}\right] \tag{B.4}$$

Then the general Bezout equation holds

$$\begin{pmatrix} \tilde{X} & -\tilde{Y} \\ -\tilde{N} & \tilde{M} \end{pmatrix} \begin{pmatrix} M & Y \\ N & X \end{pmatrix} = I \tag{B.5}$$

and $G = NM^{-1} = \tilde{M}^{-1}\tilde{N}$ are right and left coprime factorisations of G

This is equally applicable to a discrete system $G = D + C(zI - A)^1 B$.

So to apply the above to the discrete system $P_{22} = D_{22} + C_2(zI - A)^1 B_2$ MATLAB was used to calculate a suitable F and L by performing a linear quadratic regulator design

$$\min_K \sum x'x + u'u$$

for a discrete system

$$x_{k+1} = \hat{A}x_k + \hat{B}u_k$$

where $u_k = -Kx_k$. To produce the state feedback matrix $F = K$ let $\hat{A} = A, \hat{B} = B_2$ and to produce the observer gain matrix $L = K^T$ let $\hat{A} = A^T, \hat{B} = C_2^T$.

This allows the discrete state-space representations of H, U and V to be computed as follows:

$$H = \left(\begin{array}{cc|c} A - B_2F & LC_2 & LD_21 \\ 0 & A - LC_2 & B_1 + LD_{21} \\ \hline C_1 + D_{12}F & C_1 & D_{11} \end{array}\right), \tag{B.6}$$

$$U = \left(\begin{array}{c|c} A - B_2F & B_2 \\ \hline C_1 - D_{12}F & D_{12} \end{array} \right), \tag{B.7}$$

$$V = \left(\begin{array}{c|c} A - LC_2 & B_1 - LD_{21} \\ \hline C_2 & D_{21} \end{array} \right). \tag{B.8}$$

B.2 Null Chains

The interpolation conditions in theorem 3.3.2 are expressed in terms of null chains as shown below

$$(y^i_{\lambda_0} R x^j_{\lambda_0})^{(k)}(\lambda_0) = 0 \quad for \quad \begin{cases} i = 1, ..., n_u \\ j = 1, ..., n_y \\ k = 0, ..., \sigma_{U_i}(\lambda_0) + \sigma_{V_j}(\lambda_0) - 1 \end{cases}$$

where $y^i_{\lambda_0}$ are calculated from elements of the extended set of left null chains of U and $x^j_{\lambda_0}$ are calculated from elements of the extended set of right null chains of V. As mentioned in section 3.3.1 this theorem is given in terms of the discrete operator $\lambda = z^{-1}$, in which case the RHP zeros are given by the zeros λ_0 in the unit disc $\lambda_0 \in \mathcal{D}$. This expression for the interpolation conditions allow us to represent them as the linear constraints in (3.14)

$$\sum_{p=1}^{n_z} \sum_{q=1}^{n_w} \sum_{l=0}^{\infty} a^{ij,\lambda_0,k}_{pq}(l) \phi_{pq}(l) = b^{ij,\lambda_0,k}.$$

where

$$b^{ij,\lambda_0,k} = \sum_{p=1}^{n_z} \sum_{q=1}^{n_w} \sum_{l=0}^{\infty} h_{pq}(l).$$

and

$$a^{ij,\lambda_0,k}_{pq}(l) = \left[\sum_{t=0}^{\infty} \sum_{s=0}^{\infty} (y^i_{\lambda_0})^p (t - l - s)(x^j_{\lambda_0})_q(s)(\lambda^t)^{(k)} \right]_{\lambda=\lambda_0}$$

To construct these constraints we need to compute the discrete elements of the row vector $y^i_{\lambda_0}$ and the column vector $x^j_{\lambda_0}$ which are given by the left and right extended null chains respectively.

Several definitions and notations must be introduced. Right and left null chains are defined in Dahleh and Diaz-Bobillo (1995) as follows

Definition B.2.1 *Given any $m \times n$ (real) rational matrix $M(\lambda)$ analytic at λ_0, a right null chain of order σ at λ_0 is an ordered set of column vectors in $\mathcal{R}^n$, $\{x_1, x_2, ..., x_\sigma\}$, such that $x_1 \neq 0$ and*

$$T_{\lambda_0,\sigma}(M) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_\sigma \end{pmatrix} = \mathbf{0} \tag{B.9}$$

Similarly, a left null chain of order σ at λ_0 is an ordered set of column vectors in $\mathcal{R}^m$, $\{y_1, y_2, ..., y_\sigma\}$, such that $y_1 \neq 0$ and

$$T_{\lambda_0,\sigma}(M^T) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_\sigma \end{pmatrix} = \mathbf{0} \tag{B.10}$$

where $T_{\lambda_0,\sigma}(M)$ is a block-lower-triangular Toeplitz matrix:

$$T_{\lambda_0,\sigma}(M) = \begin{pmatrix} M_0 & 0 & 0 & \cdots & 0 \\ M_1 & M_0 & 0 & \cdots & 0 \\ \vdots & & & & vdots \\ M_{\sigma-1} & M_{\sigma-2} & M_{\sigma-3} & \cdots & M_0 \end{pmatrix} \tag{B.11}$$

the M_i's are given by the Taylor expansion of $M(\lambda)$ at λ_0, $M_i = \frac{1}{i!}(M)^{(i)}(\lambda_0)$.

Since $(M)^{(i)}(\lambda_0)$ is given by

$$(M)^{(i)}(\lambda_0) = \begin{cases} C_M\lambda_0(I - \lambda_0 A_M)^{-1}B_M + D_M & \text{for } i = 0 \\ i!C_M(I - \lambda_0 A_M)^{-(i+1)}A_M^{i-1}B_M & \text{for } i = 1, 2, ... \end{cases}$$

where (A_M, B_M, C_M, D_M) is the state-space representation of M, it is a straight forward matter to calculate M_i for the discrete state-space system M.

A canonical set of right null chains is defined as,

Definition B.2.2 *A canonical set of right null chains of $M(\lambda)$ at λ_0 is an ordered set of right null chains,ie, $x^i = (x_1^i, ..., x_{\sigma_i}^i)$ for $i = 1, ..., l$, such that*

1. $\{x_1^1, x_1^2, ..., x_1^l\}$ *are linearly independent,*

2. *span*$\{x_1^1, x_1^2, ..., x_1^l\} = \mathcal{N}[M(\lambda_0)]$, *and*

3. $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_l$.

and similarly for a canonical set of left null chains. An extended set of right null chains is a canonical set of right null chains augmented with $n - l$ vectors in $\mathcal{R}^n$, $\{x_1^{l+1}, x_1^{l+2}, ..., x_1^n\}$ each with order 0, such that $[x_1^1 \; x_1^2 \; ... \; x_1^n]$ is full rank.

From definition B.2.1 as long as $x_1 \neq 0$, $x_i \in \mathcal{R}^n$, then any vector in the null space of $T_{\lambda_0,\sigma}(M)$ is a candidate for the extended set of right null chains (simply transpose M for the left null chains). Therefore find the lowest σ for which the top n rows of a basis for the null space of $T_{\lambda_0,\sigma}(M)$ are all 0 and let this be σ_0. At this point we know that σ_l, the order of the highest order null chain in the extended set, is given by $\sigma_l = \sigma_0 - 1$.

Let

$$B_{\sigma_l} = \text{basis of the null space of } T_{\lambda_0,\sigma_l}(M) \tag{B.12}$$

and x = a column of B_{σ_l}. If the first j entries of x are 0, ie, $x_1 = x_2 = ... = x_j = 0$, then

$$T_{\lambda_0,\sigma_l}(M) \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ x_{j+1} \\ \vdots \\ x_{\sigma_l} \end{pmatrix} = T_{\lambda_0,\sigma_l - j}(M) \begin{pmatrix} x_{j+1} \\ x_{j+2} \\ \vdots \\ x_{\sigma_l} \end{pmatrix} = \mathbf{0} \tag{B.13}$$

with $x_{j+1} \neq 0$, in which case x with the top jn entries, 0's, removed is a right null chain of $M(\lambda)$ at λ_0 of order $\sigma_l - j$.

Therefore if we take the columns of B_{σ_l} (which is $\sigma_l n \times q$) and truncate them as described above, sort them in descending order or dimensions, ie, with σ_l first, then we have q ordered right null chains of $M(\lambda)$ at λ_0 with maximum order σ_l, ie,

$\{x^1, x^2, ..., x^q\}$. If we select the first l null chains for which the associated x_1^i are linearly independent, then we fulfill the conditions in definition B.2.2 for a canonical set of right null chains.

Now to create an extended set of right null chains the set must be extended with $n-l$ linearly independent vectors such that $[x_1^1\ x_1^2\ ...\ x_1^n]$ is full rank. This can be done using an orthogonal-triangular decomposition $X = \bar{Q}\bar{R}$, where $X = [x_1^1\ x_1^2\ ...\ x_1^l]$, $\bar{Q}$ is unitary and $\bar{R}$ is an upper triangular matrix of the same dimensions of X. Since X is $n \times l$ where $l < n$ and has l linearly independent columns then $\bar{R}$ has l linearly independent columns with the bottom $n-l$ rows all zero. This means that $\bar{R}_{1:l}$ the matrix formed by the top l rows of $\bar{R}$ is full rank. Therefore if we put $\tilde{X} = [X\ \ \bar{Q}^{l+1:n}]$ then $\tilde{X}$ is $n \times n$ and

$$\bar{Q}'[X\ \ \bar{Q}^{l+1:n}] = [\bar{R}\ \ I^{l+1:n}] = \begin{bmatrix} \bar{R}_{1:l} & 0 \\ 0 & I^{l+1:n} \end{bmatrix} \tag{B.14}$$

$$\tilde{X} = [X\ \ \bar{Q}^{l+1:n}] = \bar{Q}\begin{bmatrix} \bar{R}_{1:l} & 0 \\ 0 & I^{l+1:n} \end{bmatrix} \tag{B.15}$$

in which case $\tilde{X}$ is full rank. So $\{x^1, x^2, ..., x^n\}$ is a suitable choice for an extended set of right null chains where $\{x^1, x^2, ..., x^l\}$ are the l vectors chosen for the canonical set of right chains and $\{x^{l+1}, x^{l+2}, ..., x^n\}$ are the last $n-l$ columns of $\bar{Q}$.

So in this manner find the extended set of left nullchains of U $\{y^1, y^2, ...y^{n_z}\}$ where $y_r^i \in \mathcal{R}^{n_z}$ and the extended set of right nullchains of V $\{x^1, x^2, ...x^{n_w}\}$ where $x^j \in_l \mathcal{R}^{n_w}$. Note that U or V may need to be extended by N delays as part of the DA algorithm. Then add the vectors of the nullchain x^j as shown below to give $\hat{x}_{\lambda_0}^j(\lambda)$

$$\begin{aligned} \hat{x}_{\lambda_0}^j(\lambda) &= \sum_{l=0}^{\sigma_j-1} (\lambda-\lambda_0)^l x_{l+1}^j && \text{(B.16)} \\ &= \sum_{l=0}^{\sigma_{V_j}(\lambda_0)-1} \sum_{k=0}^{l} \begin{pmatrix} l \\ k \end{pmatrix} (-\lambda_0)^{l-k}\lambda^k x_{l+1}^j && \text{(B.17)} \\ &= \sum_{k=0}^{\sigma_{V_j}(\lambda_0)-1} c_{\lambda_0}^j(k)\lambda^k && \text{(B.18)} \end{aligned}$$

$$\text{with} \quad c^j_{\lambda_0}(k) = \sum_{l=k}^{\sigma_{V_j}(\lambda_0)-1} \begin{pmatrix} l \\ k \end{pmatrix} (\lambda_0)^{l-k} x^j_{l+1} \tag{B.19}$$

$$\tag{B.20}$$

and similarly for y^i to give $\hat{y}^i_{\lambda_0}(\lambda)$.

$$\hat{y}^i_{\lambda_0}(\lambda) = \sum_{k=0}^{\sigma_{U_i}(\lambda_0)-1} d^i_{\lambda_0}(k)\lambda^k \tag{B.21}$$

$$\text{with} \quad d^i_{\lambda_0}(k) = \sum_{r=k}^{\sigma_{U_i}(\lambda_0)-1} \begin{pmatrix} r \\ k \end{pmatrix} (\lambda_0)^{r-k} (y^i_{r+1})^T \tag{B.22}$$

If $\sigma_{V_j}(\lambda_0)$ or $\sigma_{U_i}(\lambda_0) = 0$ then the null chain is an extension chain and $\hat{x}^j_{\lambda_0}(\lambda) = x^j_1$ or $\hat{y}^i_{\lambda_0}(\lambda) = (y^i_1)^T$.

Therefore the discrete elements of the row vector $y^i_{\lambda_0}$ and the column vector $x^j_{\lambda_0}$ are given by

$$x^j_{\lambda_0}(k) = c^j_{\lambda_0}(k) \text{ for } k = 0, 1, ..\sigma_{V_j}(\lambda_0) - 1 \tag{B.23}$$

$$y^i_{\lambda_0}(k) = d^i_{\lambda_0}(k) \text{ for } k = 0, 1, ..\sigma_{U_i}(\lambda_0) - 1 \tag{B.24}$$

A MATLAB routine was implemented to compute these which takes advantage of a range of in-built and specialist MATLAB commands.

B.3 Reconditioning

For both the interpolation conditions technique and the Q-parametrisation technique a set of equality constraints are calculated on ϕ that will force it to take a suitable form such that $\Phi = H - UQV$. These are given as

$$A_{zero}\phi = b_{zero}$$

and

$$[I \quad | \quad A_q] \begin{bmatrix} \phi \\ \hline q \end{bmatrix} = h$$

respectively. These might both be expressed as a set of equality constraints

$$\tilde{A}\tilde{x} = \tilde{b} \tag{B.25}$$

There is no one unique form in which these constraints must be expressed, however for linear programming purposes it is better if the matrix $[\tilde{A} \quad \tilde{b}]$ is expressed such that it is both well conditioned and full row rank. Firstly this means that we attempt to ensure that all the significant elements of matrix are within a similar order of magnitude. It is best not to have one row expressing a constraint with coefficients in the order of 10^{-3} and another expressing a constraint with coefficients in the order of 10^3. This makes it hard for the LP solver to produce an accurate result, since the first constraint is close to 0 in comparison to the second. Secondly we do not want to repeatedly state the same constraint, just in different forms, since this increases the LP size and computational time unnecessarily. Therefore it is a good idea to recondition these equality constraints and, at the same time, remove all the inactive constraints, before proceeding any further with the building and solving of the LP. The following is based on a technique used in ℓ_1-optimal control software by I. J. Diaz-Bobillo and made available from MIT.

The method suggested to do this involves the use of the singular value decomposition, which can be computed using an in-built function of MATLAB, and is fairly simple. Find the singular decomposition of $A_b = [\tilde{A} \quad \tilde{b}]$ so that

$$A_b = \tilde{U}\tilde{S}\tilde{V}^T. \tag{B.26}$$

where $\tilde{U}$ and $\tilde{V}$ are unitary matrices, i.e., $\tilde{U}^T\tilde{U} = \tilde{U}\tilde{U}^T = I$ and the same for $\tilde{V}$. Then check the singular values σ_i (the diagonal elements of $\tilde{S}$), which decrease with increasing i, and if these values drop below the zero tolerance tol_0 of the computation then choose r so that

$$\sigma_i > tol_0, \quad \forall i \geq r. \tag{B.27}$$

Note that, if $\tilde{A}$ has n rows then, if $r < n$, A_b is not full rank and some of the equality constraints in (B.25) are inactive.

Therefore to both remove these constraints and to recondition them premultiply $\tilde{A}$ and $\tilde{b}$ by the first r rows of the transposed matrix $\tilde{U}^T$, so that

$$(\tilde{U}^T)_{1:r}\tilde{A}\tilde{x} = (\tilde{U}^T)_{1:r}\tilde{b} \quad \Leftrightarrow \quad \tilde{A}_r\tilde{x} = \tilde{b}_r. \tag{B.28}$$

Premultiply again by the diagonal matrix D^{scale} made up from the inverses of the singular values σ_i $i = 1, .., r$, i.e., $D^{scale}_{ii} = \sigma_i^{-1}$,

$$D^{scale}\tilde{A}_r\tilde{x} = D^{scale}\tilde{b}_r \quad \Leftrightarrow \quad \tilde{A}_{rs}\tilde{x} = \tilde{b}_{rs} \tag{B.29}$$

or simply let $D^{scale}_{ii} = (\max_j(\tilde{A}_r)_{ij})^{-1}$.s

Appendix C

Nonlinear Implementation

A layout for the nonlinear software implemented is shown in Figures 4.11. Implementation details for the optimisation parameters x, the constraints ($h^p(\mathrm{x})$,$g^p(\mathrm{x})$) for the feasibility subproblem (4.15) and the operating points constraints $h^{u_0}(\mathrm{x})$, are given in the following appendices.

C.1 The optimisation parameters

The optimisation parameters x should be constructed, so that

$$\mathrm{x}(ind(p,j,k)) = scale(u_j^p(k)) \tag{C.1}$$

where $ind(p,j,k)$ is a function giving the indices in x of the scaled discrete control input elements $u_j^p(k)$. $scale$ is a function which scales all your $u_j^p(k)$ so that they have the same range, i.e., we use

$$scale(u_j^p(k)) = \frac{(u_j^p(k) - u^l)}{u^h - u^l} + 0.5 \tag{C.2}$$

so that all $\mathrm{x}(i)$'s range as follows

$$0.5 \leq \mathrm{x} \leq 1.5$$

C.2 The feasibility subproblem constraints

The feasibility subproblem (4.15) for $w^p \in \bar{W}$ provides both equality and inequality constraints, i.e.,

$$h^p(\mathrm{x}) = 0 \Leftrightarrow \begin{cases} \dot{\mathrm{x}}^p(t_0) = 0 \\ F(\dot{x}^p, x^p, y^p, u^p, w^p, t) = 0 \end{cases}$$

and

$$g^p(\mathrm{x}) \geq 0 \Leftrightarrow -c(x^p, y^p, u^p) \geq 0$$

where $u^p(0) = u_0$ has been substituted to drop u_0. The equality constraints $h^p(\mathrm{x})$ describe the system and are, in fact, evaluated by the DAE solver DASOLV to give $\dot{x}^p(t)$, $x^p(t)$ and $y^p(t)$, for a series of discrete times, for given u^p and w^p. These values are used to evaluate the performance constraints in $g^p(\mathrm{x})$

$$-c(x^p, y^p, u^p) \geq 0 \tag{C.3}$$

and their gradients $\nabla g^p(\mathrm{x})$

$$-\frac{\partial c(x^p, y^p, u^p)}{\partial \mathrm{x}} \tag{C.4}$$

The user must provide not only the function $C = c(x^p, y^p, u^p)$ in terms of the variables x^p, y^p and u^p, but also the gradient, i.e.,

$$\frac{\partial C}{\partial \mathrm{x}} = \frac{\partial C}{\partial x^p}\frac{\partial x^p}{\partial \mathrm{x}} + \frac{\partial C}{\partial y^p}\frac{\partial y^p}{\partial \mathrm{x}} + \frac{\partial C}{\partial u^p}\frac{\partial u^p}{\partial \mathrm{x}} \tag{C.5}$$

where the $\frac{\partial C}{\partial x^p}$,$\frac{\partial C}{\partial y^p}$ and $\frac{\partial C}{\partial u^p}$ are user supplied functions of x^p, y^p and u^p, but the sensitivities $\frac{\partial x^p}{\partial \mathrm{X}}$ and $\frac{\partial y^p}{\partial \mathrm{X}}$ are provided by DASOLV. The sensitivity of the control inputs to the optimisation parameters are given by

$$\frac{\partial u^p}{\partial \mathrm{x}(i)} = \begin{cases} u^h - u^l, & i = ind(p, j, k) \\ 0, & \text{otherwise} \end{cases} \tag{C.6}$$

where $ind(p, j, k)$ is described in section C.1.

C.3 The operating points constraints

To ensure a common operating point for each feasibility subproblem given by $w^p \in \bar{W}$, a set of $n_u(n_{dist} - 1)$ extra equality constraints, $h^{u_0}(\mathrm{x})$ are added to the problem. These constraints enforce the following conditions.

$$u_j^p(0) - u_j^{p-1}(0) = 0, \quad \text{for} \begin{cases} p = 2, ...n_{dist} \\ j = 1, ...n_u \end{cases}$$

Therefore the gradients $\nabla h^{u_0}(\mathrm{x})$ must be given by the user as shown below

$$\frac{\partial h^{u_0}(\mathrm{x})}{\partial \mathrm{x}(i)} = \begin{cases} u_j^h - u_j^l, \quad i = ind(p, j, k) \\ -(u_j^h - u_j^l), \quad i = ind(p-1, j, k) \\ 0, \quad \text{otherwise} \end{cases} \tag{C.7}$$

Also the objective function's gradient

$$\nabla f(\mathrm{x}) = \frac{\partial f(\mathrm{x})}{\partial \mathrm{x}} \tag{C.8}$$

should be user supplied in a similar manner to (C.5) and will be problem specific.

Appendix D

Details of Case Studies

D.1 The Linearised Reduced Reactor Model

The linearised reduced continuous state space model of the adiabatic PFR's in 5.3 is described by

$$\begin{aligned} \dot{x} &= A\ x + [B_1\ \ B_2] \begin{bmatrix} w \\ u \end{bmatrix} \\ y &= C\ x + [D_1\ \ D_2] \begin{bmatrix} w \\ u \end{bmatrix} \end{aligned} \tag{D.1}$$

where A is 24×24 and is given by

Columns 1 through 6

$-9.1026e4$	$-3.7991e-8$	$-5.0529e-9$	$2.1778e-10$	$-1.7125e-9$	$5.7078e-9$
$7.7201e-9$	$-7.9508e4$	$7.9814e-10$	$-1.4387e-10$	$1.4608e-9$	$-2.1827e-9$
$1.5352e-9$	$-3.8298e-11$	$-2.2422e3$	$5.8813e3$	$-1.6594e-9$	$7.2745e-10$
$7.1723e-9$	$1.2394e-8$	$-5.8813e3$	$-2.2422e3$	$1.5780e-9$	$-5.4971e-11$
$-1.8497e-8$	$-3.2556e-8$	$-4.3311e-9$	$5.3441e-10$	$-2.2832e3$	$3.6156e3$
$3.0815e-8$	$4.9772e-8$	$3.5998e-9$	$-4.2909e-9$	$-3.6156e3$	$-2.2832e3$
$1.2036e-9$	$5.4978e-9$	$4.2775e-10$	$-3.7559e-10$	$-3.7369e-10$	$-1.0114e-9$
$-1.7698e-9$	$-3.3735e-9$	$-2.0997e-10$	$1.2211e-10$	$-7.7008e-10$	$-6.8110e-10$
$8.2543e-8$	$1.3267e-7$	$9.3404e-9$	$-2.0299e-8$	$-3.7115e-9$	$-8.5131e-9$
$-1.7188e-8$	$-3.4521e-8$	$-1.1983e-9$	$8.2559e-9$	$1.1509e-8$	$3.4062e-9$

Columns 1 through 6 (continued)

$-1.7188e-8$	$-3.4521e-8$	$-1.1983e-9$	$8.2559e-9$	$1.1509e-8$	$3.4062e-9$
$1.3044e-7$	$2.1559e-7$	$1.7251e-8$	$-4.5290e-8$	$-2.5807e-8$	$-3.0373e-8$
$1.0184e-7$	$1.5979e-7$	$1.5232e-8$	$-3.2115e-8$	$-1.2238e-8$	$-2.2515e-8$
$-4.4193e-10$	$6.0887e-10$	$-7.7236e-12$	$-6.8623e-10$	$2.8266e-10$	$2.1502e-10$
$2.6626e-9$	$4.1150e-9$	$3.2934e-10$	$5.4051e-10$	$6.6206e-10$	$7.1412e-10$
$-8.6101e-9$	$-1.6994e-8$	$-1.2614e-9$	$5.2624e-10$	$-1.6288e-9$	$1.0418e-9$
$-9.5584e-10$	$-3.9241e-9$	$-1.2158e-9$	$-2.3648e-9$	$1.7386e-9$	$1.8536e-9$
$2.9470e-9$	$7.1547e-9$	$4.3607e-10$	$2.0602e-9$	$-1.3110e-9$	$-1.1704e-9$
$-6.4833e-9$	$-3.1868e-8$	$-8.5780e-9$	$-1.3925e-9$	$-1.1412e-9$	$6.9431e-9$
$2.1283e-8$	$4.1078e-8$	$2.1216e-9$	$5.9768e-10$	$1.1154e-9$	$-4.8177e-9$
$-2.5067e-8$	$-4.8252e-8$	$2.6325e-10$	$5.9789e-9$	$-3.7073e-9$	$3.6685e-9$
$-6.2931e-9$	$-3.3226e-8$	$-9.6135e-9$	$-3.0497e-9$	$1.5672e-9$	$8.2480e-9$
$9.7301e-9$	$8.6760e-8$	$2.8205e-8$	$6.0291e-9$	$4.1413e-10$	$-2.4165e-8$
$-2.5011e-8$	$-5.0105e-8$	$-8.8515e-10$	$3.9354e-9$	$-2.0017e-9$	$5.4381e-9$
$-9.5571e-9$	$-7.5511e-8$	$-2.3410e-8$	$-5.3474e-9$	$1.9316e-10$	$2.0747e-8$

Columns 7 through 12

$-9.9080e9$	$-2.1709e-10$	$-2.2234e-9$	$1.0006e-8$	$9.1346e-9$	$-9.5475e-9$
$5.4979e-9$	$2.6174e-9$	$1.1737e-9$	$-3.8059e-9$	$-3.5418e-9$	$3.5663e-9$
$-5.7476e-9$	$-2.6402e-9$	$-6.7542e-10$	$6.6304e-10$	$8.3066e-10$	$-6.7256e-10$
$1.4070e-9$	$2.6912e-9$	$7.0967e-10$	$2.1436e-9$	$1.4488e-9$	$-1.7569e-9$
$-1.8509e-8$	$-1.2360e-8$	$-3.3627e-9$	$-3.9717e-9$	$-2.2497e-9$	$3.5119e-9$
$-8.0906e-9$	$1.3785e-9$	$6.3880e-10$	$1.1040e-8$	$9.2131e-9$	$-1.0138e-8$
$-9.3614e2$	$3.8932e3$	$5.2684e-10$	$-1.6266e-9$	$-1.1674e-9$	$9.9634e-10$
$-3.8932e3$	$-9.3614e2$	$1.7067e-10$	$-2.1419e-9$	$-1.5780e-9$	$1.5780e-9$
$1.0626e-8$	$1.5359e-8$	$-2.3600e3$	$1.9208e3$	$1.8584e-8$	$-2.3014e-8$
$2.6246e-8$	$7.8485e-9$	$-1.9208e3$	$-2.3600e3$	$9.1776e-10$	$-8.6829e-10$
$4.2819e-9$	$2.1332e-8$	$1.9162e-8$	$7.9388e-9$	$-2.5408e3$	$5.3850e2$
$2.6276e-8$	$2.3840e-8$	$1.6231e-8$	$1.4636e-8$	$-5.3850e2$	$-2.5408e3$
$1.2233e-9$	$-3.2456e-10$	$-1.4025e-10$	$1.1224e-9$	$8.7724e-10$	$-8.2168e-10$
$-4.9670e-9$	$-2.0545e-9$	$-2.0986e-10$	$9.9807e-10$	$5.5933e-10$	$-4.6684e-10$

Columns 7 through 12 (continued)

$-9.3891e-$	$1.8967e-9$	$-4.2106e-10$	$1.0835e-9$	$1.4580e-9$	$-1.7039e-9$
$-3.5185e-9$	$-7.9554e-9$	$-1.0248e-9$	$5.7802e-9$	$4.4627e-9$	$-4.1649e-9$
$2.6692e-9$	$8.8205e-9$	$7.8312e-10$	$-3.9101e-9$	$-2.8731e-9$	$2.5141e-9$
$-9.4453e-9$	$2.4993e-9$	$-2.5798e-9$	$1.1493e-8$	$1.0532e-8$	$-1.1010e-8$
$7.7890e-9$	$8.1611e-9$	$2.8715e-9$	$-9.5588e-9$	$-8.2851e-9$	$7.8869e-9$
$-4.1036e-9$	$4.2082e-9$	$-2.5426e-9$	$2.1231e-9$	$2.1808e-9$	$-1.6619e-9$
$-1.1144e-8$	$-6.5483e-9$	$-3.7249e-9$	$1.5505e-8$	$1.3110e-8$	$-1.2810e-8$
$3.3179e-8$	$-3.8398e-9$	$9.3847e-9$	$-4.1689e-8$	$-3.6776e-8$	$3.7543e-8$
$-7.6117e-9$	$-2.8610e-9$	$-3.4854e-9$	$7.4038e-9$	$6.2233e-9$	$-5.4239e-9$
$-2.9153e-8$	$4.5743e-10$	$-8.2317e-9$	$3.6543e-8$	$3.1952e-8$	$-3.2415e-8$

Columns 13 through 18

$1.5109e-8$	$-9.1736e-9$	$6.4946e-9$	$-7.5048e-9$	$2.3081e-9$	$-2.4081e-9$
$-1.0332e-8$	$2.8303e-9$	$-2.4538e-9$	$5.0036e-9$	$-6.1814e-10$	$1.2678e-9$
$8.0117e-9$	$-9.0631e-10$	$2.1066e-9$	$-4.9329e-9$	$-2.5887e-9$	$-9.7441e-10$
$-4.3980e-9$	$-4.0899e-10$	$3.8460e-9$	$1.6089e-9$	$-4.5677e-9$	$-2.3428e-9$
$3.9692e-8$	$-6.6771e-9$	$-1.4318e-8$	$-1.6912e-8$	$1.1083e-8$	$8.3161e-9$
$3.0828e-9$	$-3.7527e-9$	$1.4113e-8$	$-7.9527e-9$	$-2.4207e-8$	$-1.0198e-8$
$-1.6055e-8$	$3.1034e-9$	$1.3059e-9$	$7.9660e-9$	$1.9212e-9$	$1.5922e-9$
$-1.8305e-9$	$1.4794e-9$	$-5.0802e-10$	$-3.4089e-10$	$9.3392e-10$	$1.7865e-9$
$-5.2938e-8$	$1.5698e-8$	$5.5716e-8$	$-2.2174e-8$	$-7.4815e-8$	$-1.3067e-8$
$-4.3608e-8$	$4.1207e-9$	$-4.7274e-10$	$4.0309e-8$	$2.7953e-8$	$-1.6407e-8$
$-5.4493e-8$	$3.7575e-8$	$1.0026e-7$	$-1.0578e-7$	$-1.5492e-7$	$2.6240e-8$
$-8.5293e-8$	$3.6608e-8$	$8.6767e-8$	$-5.5622e-8$	$-1.1432e-7$	$4.4882e-9$
$-1.0602e3$	$1.1749e3$	$3.9318e-10$	$1.1848e-9$	$6.4987e-10$	$-1.2815e-9$
$-1.1749e3$	$-1.0602e3$	$-1.0629e-9$	$-3.4875e-9$	$-1.6560e-9$	$-1.5913e-9$
$1.7750e-9$	$-1.0712e-9$	$-8.7677e1$	$1.5650e2$	$2.1316e-9$	$1.7661e-9$
$8.5099e-9$	$-1.8405e-9$	$-1.5650e2$	$-8.7677e1$	$-3.9414e-10$	$-5.7731e-9$
$-8.5964e-9$	$-2.3315e-10$	$2.8675e-9$	$3.2456e-9$	$-9.1484e1$	$1.5968e2$
$8.1044e-9$	$-1.0885e-8$	$4.0902e-9$	$-5.1303e-9$	$-1.5968e2$	$-9.1484e1$
$-1.9958e-8$	$5.0556e-9$	$-4.9868e-9$	$1.1275e-8$	$-2.7988e-9$	$5.9401e-9$

Columns 13 through 18 (continued)

$1.2283e-8$	$-4.5878e-9$	$7.3943e-9$	$-1.3116e-9$	$2.5398e-9$	$-9.1303e-10$
$1.7208e-8$	$-1.1133e-8$	$1.1343e-10$	$-6.3258e-9$	$-2.9577e-10$	$-6.6258e-9$
$-2.9073e-8$	$3.9520e-8$	$-6.8963e-9$	$9.3263e-9$	$-5.1268e-9$	$9.4120e-9$
$2.0392e-8$	$-6.4283e-9$	$5.5711e-9$	$-4.9931e-9$	$1.5102e-9$	$-6.0651e-9$
$2.8117e-8$	$-3.3947e-8$	$5.0406e-9$	$-7.8233e-9$	$4.9359e-9$	$-1.0437e-8$

Columns 19 through 24

$1.1045e-8$	$-1.4897e-8$	$6.0338e-9$	$-5.3692e-9$	$1.5880e-8$	$-4.7889e-9$
$-4.2078e-9$	$5.0090e-9$	$-3.2916e-9$	$3.8252e-9$	$-5.0226e-9$	$3.4494e-9$
$1.9907e-9$	$-3.0553e-9$	$1.8774e-10$	$-4.6283e-9$	$2.6605e-9$	$-5.1544e-9$
$1.1170e-9$	$-9.2585e-10$	$-1.5204e-9$	$-6.2190e-9$	$1.7492e-9$	$-7.4312e-9$
$-1.4725e-9$	$2.3056e-9$	$6.4473e-9$	$1.5019e-8$	$-6.0935e-9$	$1.7517e-8$
$4.3435e-9$	$-4.9370e-9$	$-5.1390e-9$	$-3.1151e-8$	$5.7079e-9$	$-3.6837e-8$
$-7.6585e-10$	$-1.1693e-10$	$-7.3008e-10$	$4.9212e-9$	$9.4637e-10$	$4.9575e-9$
$-2.9785e-10$	$-1.5026e-9$	$1.1354e-11$	$2.5274e-9$	$1.1750e-9$	$2.7832e-9$
$1.5253e-8$	$-5.3805e-8$	$-1.3860e-8$	$-9.1386e-8$	$5.7903e-8$	$-1.1203e-7$
$-5.7603e-9$	$2.1148e-8$	$-1.9602e-10$	$2.5263e-8$	$-1.7940e-8$	$3.4851e-8$
$3.9400e-8$	$-1.6062e-7$	$-1.9214e-8$	$-1.7082e-7$	$1.6421e-7$	$-2.2096e-7$
$2.7742e-8$	$-1.1538e-7$	$-1.7011e-8$	$-1.3236e-7$	$1.2055e-7$	$-1.6768e-7$
$-2.0984e-10$	$4.6221e-10$	$9.0738e-10$	$-8.4872e-10$	$-6.3594e-13$	$-9.0616e-10$
$-8.2423e-13$	$1.7992e-9$	$-7.4624e-10$	$-2.6809e-9$	$-2.2221e-9$	$-2.7382e-9$
$4.2428e-9$	$-9.0810e-9$	$3.2747e-9$	$4.0241e-10$	$9.7747e-9$	$4.8031e-10$
$-1.1959e-9$	$3.9843e-9$	$2.7854e-9$	$-7.8863e-9$	$-4.2238e-9$	$-8.3414e-9$
$2.0471e-9$	$-6.0660e-9$	$-3.6109e-9$	$8.0502e-9$	$6.6241e-9$	$8.3202e-9$
$1.0382e-8$	$-1.6486e-8$	$2.7724e-9$	$-3.4220e-9$	$1.6758e-8$	$-3.5070e-9$
$-7.2161e1$	$6.8310e1$	$-1.0484e-8$	$1.1343e-8$	$-1.0071e-8$	$1.0776e-8$
$-6.8310e1$	$-7.2161e1$	$6.2197e-9$	$-1.5788e-9$	$1.0647e-8$	$1.8628e-9$
$7.5201e-9$	$-5.7417e-9$	$-7.6985e1$	$6.9110e1$	$5.8257e-9$	$-1.0662e-8$
$-3.0480e-8$	$3.8834e-8$	$-6.9110e1$	$-7.6985e1$	$-3.9848e-8$	$1.3794e-8$
$9.3759e-9$	$-4.5090e-9$	$8.1337e-9$	$-8.8186e-9$	$-6.7752e1$	$-5.7825e-9$
$2.5343e-8$	$-2.9816e-8$	$6.7777e-9$	$-1.3871e-8$	$3.0781e-8$	$-7.3262e1$

B_1 and B_2 are both 24×2 and $[B_1 \;\; B_2]$ is given by

$-6.6969e2$	$-5.4179e2$	$4.4311e1$	1.1900
$3.1984e2$	4.6564	$-3.9364e1$	-1.4925
-2.8428	$8.0554e-1$	2.1869	5.4068
$-4.8254e-$	$17.9315e-2$	-2.7225	-1.8381
-5.4129	$7.1837e-1$	$-7.6219e-1$	$1.8988e1$
2.6885	$-2.7035e-1$	-9.5522	$-1.5505e1$
$8.7455e-1$	$-9.5707e-2$	$-3.4516e-1$	-1.3710
$3.3820e-1$	$-4.0559e-1$	2.2015	$4.4571e-1$
4.1806	$8.8613e-1$	$1.6915e1$	$-6.3643e1$
-6.0760	1.8070	$1.8976e1$	$1.6809e1$
-5.1905	3.5129	$4.1284e1$	$-9.8800e1$
$-1.2836e1$	5.6625	$5.8074e1$	$-7.0754e1$
$-9.0597e-1$	$-2.1648e-1$	1.1047	1.8060
$-8.4346e-1$	$-1.7710e-1$	-3.1062	-1.5664
$-4.2669e2$	$-2.4561e1$	6.5573	2.1068
$1.7767e2$	$-7.7590e1$	-5.1685	$1.1543e1$
$-1.1969e2$	$-8.5803e1$	4.2686	-6.8490
$-2.2224e2$	$-1.0741e3$	3.7405	1.8707
$8.4893e2$	$-6.1536e1$	-8.7353	-2.7574
$-1.4075e3$	$1.2333e2$	$1.4884e1$	$-5.1282e1$
$-3.1449e1$	$-1.0347e3$	-2.3112	5.7209
$3.1634e2$	$3.7314e3$	-7.4159	7.8021
$-1.2241e3$	$3.8248e1$	$1.0899e1$	$-4.1061e1$
$-3.1096e2$	$-3.0888e3$	7.6702	-9.1807

C is 4×24 and is given by

Columns 1 through 6

$8.0025e4$	$3.1945e4$	$-6.1877e2$	$-1.5772e3$	$-3.3619e3$	$2.0570e3$
$-5.1658e-1$	$-2.2538e-1$	$-1.5065e-1$	$-1.8573e-1$	$-7.5234e-1$	$2.8808e-1$
$-7.2095e-3$	$-1.2739e-2$	$-3.0802e-3$	$-2.1603e-3$	$-2.2132e-3$	$6.7502e-4$
$4.2700e-3$	$7.9424e-3$	$-1.3461e-3$	$4.1355e-4$	$-1.4176e-4$	$6.8714e-4$

Columns 7 through 12

$-7.8116e2$	$1.3076e2$	$-3.0411e3$	$-3.8347e3$	$-2.8862e3$	$4.4349e3$
$-1.2860e2$	$-2.1614e1$	$-4.2880e-1$	$-6.1328e-1$	$-3.3563e-1$	$5.5412e-1$
$3.8000e-2$	$-4.5109e-2$	$5.0004e-4$	$-1.1446e-3$	$-2.3896e-4$	$-5.1152e-4$
$2.5236e-2$	$-2.8204e-2$	$-3.2261e-4$	$-3.5679e-4$	$-2.8861e-4$	$3.6287e-4$

Columns 13 through 18

$4.4671e2$	$1.0190e3$	$-8.5058e-1$	-1.4674	$7.2693e-1$	$4.2562e-1$
$-1.8943e2$	$-1.1715e2$	$9.8667e-2$	$-1.7612e-1$	$-7.1644e-3$	$4.3546e-2$
$7.0669e-2$	$7.6391e-2$	$2.9974e-2$	$1.6913e-1$	$-1.4770e-3$	$-1.2853e-2$
$1.4610e-2$	$2.4364e-2$	$2.2070e-2$	$-3.1052e-2$	$-6.6635e-2$	$-2.6015e-2$

Columns 19 through 24

$8.0939e-2$	$-9.5773e-1$	$2.2128e-1$	$9.7850e-1$	1.6490	$9.2049e-1$
$2.0318e-1$	$-2.1551e-1$	$-4.6984e-2$	$7.6519e-2$	$4.2868e-1$	$1.1673e-1$
$1.1909e-2$	$2.1346e-1$	$-3.0267e-3$	$-8.8313e-3$	$-2.2218e-1$	$-4.0983e-3$
$5.9647e-3$	$-1.3476e-2$	$-1.9336e-2$	$-9.6279e-2$	$1.9373e-2$	$-9.8857e-2$

D_1 and D_2 are both 4×2 and $[D_1 \ \ D_2]$ is given by

$4.6888e2$	$4.8517e2$	$-1.8877e1$	$-1.6796e-1$
$-1.2078e-3$	$-1.0149e-3$	$-5.7959e-4$	$2.0480e-6$
$-2.9297e-5$	$-2.6383e-5$	$-1.3517e-5$	$4.9146e-8$
$-9.7889e-6$	$-9.7834e-6$	$-4.2906e-6$	$1.6211e-8$

(Note in the above $-9.7834e-6$ means -9.7834×10^{-6} and so on.)

The variable vectors are given by

$$y = \begin{bmatrix} fB \\ lc_{1,out} \\ Tout1 \\ Tout2 \end{bmatrix}, u = \begin{bmatrix} Tin1 \\ Tin2 \end{bmatrix}, w = \begin{bmatrix} F \\ c_{1,in} \end{bmatrix}$$

and the model has been scaled so that:

- $-1 \leq w \leq 1$,
- u has a range of 1, i.e., $-0.6263 \leq Tin1 \leq 0.3737$ and $0 \leq Tin2 \leq 1$,
- $y(2:4)$ is given by, $-\infty \leq lc_{1,out} \leq 1$, $-1 \leq Tout1 \leq 0$ and $-0.04654 \leq Tout2 \leq 0.95346$.

$y(1)$ is the objective and is not scaled.

D.2 The Nonlinear Evaporator Model

The variables are described as follows:

- F1 is the feed flowrate (kg/min)
- F2 is the product flowrate (kg/min)
- F3 is the circulating flowrate (kg/min)
- F4 is the vapor flowrate (kg/min)
- F5 is the condensate flowrate (kg/min)
- X1 is the feed composition (%)
- X2 is the product composition (%)
- T1 is the feed temperature (° C)

- T2 is the product temperature (° C)
- T3 is the vapor temperature (° C)
- L2 is the separator level (m)
- P2 is the operating pressure (kPa)
- F100 is the steam flowrate (kg/min)
- T100 is the steam temperature (° C)
- P100 is the steam pressure (kPa)
- Q100 is the heater duty (kW)
- F200 is the cooling water flowrate (kg/min)
- T200 is the cooling water inlet temperature (° C)
- T201 is the cooling water outlet temperature (° C)
- Q200 is the condenser duty (kW)

The equations are given as:

- Process liquid mass balance:

  $\mathrm{pA}\ \dot{L}2 = \mathrm{F1 - F4 - F2}$

- Process liquid solute mass balance:

  $\mathrm{M}\ \dot{X}2 = (\mathrm{F1\ X1}) - (\mathrm{F2\ X2})$

- Process vapour mass balance:

  $\mathrm{C}\ \dot{P}2 = \mathrm{F4 - F5}$

- Process liquid energy balance:

  $\mathrm{T2} = (0.5616\ \mathrm{P2}) + (0.3126\ \mathrm{X2}) + 48.43$

  $\mathrm{T3} = (0.507\ \mathrm{P2}) + 55.0$

  $\mathrm{F4} = (\mathrm{Q100} - (\mathrm{F1\ Cp\ (T2 - T1)}))/\lambda$

- Heat steam jacket:

  T100 = (0.1538 P100) + 90.0

  Q100 = 0.16 (F1 + F3) (T100 - T2)

  F100 = Q100 / λ_s

- Condenser:

  Q200 = (UA2 (T3 - T200)) / (1 + (UA2 / (2 Cp F200)))

  T201 = T200 + (Q200 / (F200 Cp))

  F5 = Q200 / λ

where the constants are described as:

- pA, (liquid density) $\times$ (cross-sectional area of the separator),=20 kg/m
- M, amount of liquid in evaporator,=20 kg
- C, constant that converts the mass of vapor into an equivalent pressure,=4 kg/kPa
- Cp, heat capacity of liquor,=0.07 kW/K(kg/min)
- λ, latent heat of vaporisation of liquor,=38.5 kW/(kg/min)
- λ_s, latent heat of steam,=36.6 kW/(kg/min)
- UA2, (overall heat transfer coefficient) $\times$ (the heat transfer area),=6.84 kW/K.

The steady state levels of the four possible disturbances are $F1_{ss}$=10 kg/min, $X1_{ss}$=5.0 %, $T1_{ss}$=40 ° C, $T200_{ss}$=25 ° C, while the steady state values of F2, P100, F200 and F3 are chosen either for the level PID or by the optimisation.

D.3 Path Constraints on X2 and P2

The path constraints

$$\begin{aligned} -\frac{range(\%)}{100\%}X2_{ss} \leq \quad \Delta X2 \quad &\leq \frac{range(\%)}{100\%}X2_{ss} \\ -\frac{range(\%)}{100\%}P2_{ss} \leq \quad \Delta P2 \quad &\leq \frac{range(\%)}{100\%}P2_{ss} \end{aligned} \tag{D.2}$$

can be enforced by creating new variables in the model and putting final time constraints on these into the optimisation.

For example if we have the path constraint:

$$y_l \leq y \leq y_u$$

for some variable y, then we create the new variables δy_l and δy_u as follows

$$\begin{aligned}
\delta y_l &= tol_y \int_0^{t_f} [\max(y_l - y, 0)]^2.dt \\
\delta y_u &= tol_y \int_0^{t_f} [\max(y - y_u, 0)]^2.dt
\end{aligned}$$

and put the inequality constraints

$$\begin{aligned}
optacc(1 - \delta y_l) &\geq 0 \\
optacc(1 - \delta y_u) &\geq 0
\end{aligned}$$

in the optimisation. Here *optacc* is the optimisation accuracy used and was selected to be 10^{-5} in this example, while $tol_y \geq 0$ should be chosen so that the violations

$$\delta y_l \text{ or } \delta y_u = tol_y$$

are acceptable.

D.4 The Linearised Evaporator Model

The linearised continuous state space model of the evaporator is described by

$$\begin{aligned}
\dot{x} &= A \quad x \quad + \quad [B_1 \;\; B_2] \begin{bmatrix} w \\ u \end{bmatrix} \\
y &= C \quad x \quad + \quad [D_1 \;\; D_2] \begin{bmatrix} w \\ u \end{bmatrix}
\end{aligned} \tag{D.3}$$

where

$$A = \begin{bmatrix} -1.5402 \times 10^{-1} & -1.5718 \times 10^{-1} \\ -6.6706 \times 10^{-3} & -3.9477 \times 10^{-2} \end{bmatrix},$$

$$B_1 = \begin{bmatrix} -2.8472 \times 10^{-1} & 1.0000 \times 10^{-1} \\ 1.4236 \times 10^{-2} & 0 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 3.6931 \times 10^{-1} & 0 & 2.8472 \times 10^{-1} \\ 4.5605 \times 10^{-2} & -5.3764 \times 10^{-3} & 3.5159 \times 10^{-2} \end{bmatrix},$$

$C = I_2$, $D_1 = 0_{2\times 2}$ and $D_2 = 0_{2\times 3}$. The vectors w, u, x and y are given as follows

$$w = \begin{bmatrix} \mathrm{F1} \\ \mathrm{X1} \end{bmatrix}, u = \begin{bmatrix} \mathrm{P100} \\ \mathrm{F200} \\ \mathrm{F3} \end{bmatrix}, x = \begin{bmatrix} \mathrm{X2} \\ \mathrm{P2} \end{bmatrix},$$

and $y = x$. This model has been scaled by 10% deviations in all the variables.

References

H. Ajbar, M. R. Keenan, and J. C. Kantor. Optimal linear regulation with hard constraints. *AIChE Journal*, 41(11):2439–2450, November 1995.

B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Prentice-Hall, Inc., 1989.

M. Athans and P. L. Falb. *Optimal Control*. McGraw-Hill, New York, 1966.

S. P. Boyd and C. H. Barratt. *Linear controller design : limits of performance*. Prentice-Hall, Inc, 1991.

S. P. Boyd, V. Balakrishnan, C. H. Barratt, N. M. Khraishi, X. Li, D. G. Meyer, and S. A. Norman. A new CAD method and associated architectures for linear controllers. *IEEE Transactions on Automatic Control*, 33(3):268–282, March 1988.

S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control theory*. SIAM, Philadelpia, 1994.

A. .E. Bryson and Y. C. Ho. *Applied Optimal Control*. Hemispere Publishing, 1975.

C. L. Chen and S. Macchietto. SRQPD- Version 1.1: Users Guide. Technical report, Centre for Process Systems Engineering, Imperial College, 1989.

S. Chenery and S. Walsh. ℓ_1-optimisation and process controllability analysis. In *UKACC International Conference on CONTROL '96*, volume 1, pages 503–508, September 1996.

S. D. Chenery and S. P. Walsh. Process controllability analysis using linear programming. *Accepted for publication in Journal of Process Control*, 1997.

CPLEX Optimization, Inc. *Using the CPLEX callable Library.*

M. A. Dahleh and I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach.* Prentice-Hall, Inc, 1995.

M. A. Dahleh and J. B. Pearson. l^1 - Optimal-Feedback Controllers for Discrete-Time Systems. In *Proceedings of American Control Conference, June 18-20, 1986*, volume 3, pages 1964–1968, June 1986.

M. A. Dahleh and J. B. Pearson. l^1 - Optimal Compensators for Continuous-Time Systems. *IEEE Transactions on Automatic Control*, AC-32(10):889–895, October 1987.

M. A. Dahleh and J. B. Pearson. l^1 - Optimal Feedback Controllers for MIMO Discrete-Time Systems. *IEEE Transactions on Automatic Control*, AC-32(4):314–322, April 1987.

P. Daoutidis and C. Kravaris. Inversion and zero dynamics in nonlinear multivaribale control. *AIChE Journal*, 37(4):527–538, April 1991.

P. Daoutidis and C. Kravaris. Structural evaluation of control configurations for multivariable nonlinear processes. *Chemical Engineering Science*, 47(5):1091–1107, 1992.

V. D. Dimitriadis and E. N. Pistikopoulos. Flexibilty analysis of dynamic systems. *Industrial and Engineering Chemistry Research*, 34(12):4451–4462, 1995.

T. F. Edgar and D. M. Himmelblau. *Optimization of Chemical Processes.* McGraw-Hill, New York, 1988.

O. Frank. Economic evaluation of achievable control performance using linear programming. Diploma Thesis, Centre Report Series No. B97.33, Centre for Process Systems Engineering, Imperial College, 1997.

M. Green and D. J. N. Limebeer. *Linear Robust Control*. Prentice-Hall, Inc, 1995.

I. E. Grossman, K. P. Haleman, and R. E. Swaney. Optimization strategies for flexible chemical processes. *Computers and Chemical Engineering*, 7(4):439–462, 1983.

R. M. Hirschorn. Invertibility of Multivariable Nonlinear Control Systems. *IEEE Transactions on Automatic Control*, AC-24(6):855–865, December 1979.

B. R. Holt and M. Morari. Design of Resilient Processing Plants-V. The Effect of Deadtime on Dynamic Resilience. *Chemical Engineering Science*, 40(7):1229–1237, 1985.

R. B. Jarvis and C. C. Pantelides. DASOLV - A Differential-Algebraic Equation Solver. Technical report, Centre for Process Systems Engineering, Imperial College, 1992.

F. L. Lewis. *Optimal Control*. John Wiley and Sons, 1986.

C. W. Li and Y .K. Feng. Functional Reproducibility of General Multivariable Analytic Nonlinear Systems. *International Journal of Control*, 45(1):255–268, 1987.

V. Manousiouthakis and M. Nikolaou. Analysis of decentralised control structures for nonlinear systems. *AIChE Journal*, 35(4):549–558, April 1989.

G. Mijares, C. D. Holland, R. McDaniel, C. R. Dollar, and S. E. Gallun. Analysis and evaluation of the relative gains for nonlinear systems. *Computers and Chemical Engineering*, 9(1):61–70, 1985.

M. J. Mohideen, J. D. Perkins, and E. N. Postikopoulos. Optimal design of dynamic systems under uncertainty. *AIChE Journal*, 42(8):2251–2272, August 1996.

M. Morari and E. Zafiriou. *Robust Process Control*. Prentice-Hall, Inc, 1989.

M. Morari, E. Zafiriou, and B. R. Holt. Design of Resilient Processing Plants. New Characterization of the Effect of RHP Zeros. *Chemical Engineering Science*, 42(10):2425–2427, 1987.

M. Morari. Design of Resilient Processing Plants-III. A General Framework for the Assessment of Dynamic Resilience. *Chemical Engineering Science*, 38(11):1881–1891, 1983.

M. Morari. Effect of design on the controllability of chemical plants. In *Interactions between Process Design and Process Control;Ed. J. D. Perkins*, pages 3–16. Pergamon, Oxford, 1992.

L. T. Narraway, J. D. Perkins, and G. W. Barton. Interaction between process design and process control: economic analysis of process dynamics. *Journal of Process Control*, 1:243–250, November 1991.

R. B. Newell and P. L. Lee. *Applied Process Control: A Case Study*. Prentice-Hall of Australia Ltd, 1989.

A. V. Oppenheim, A. S. Willsky, and I. T. Young. *Signals and Systems*. Prentice-Hall, Inc., 1983.

J. D. Perkins and S. P. K. Walsh. Optimisation as a tool for design/control integration. *Computers and Chemical Engineering*, 20:315–323, 1996.

J. D. Perkins and M. P. F. Wong. Assessing Controllability of Chemical Plants. *Chemical Engineering Research and Design*, 63(6):358–362, 1985.

P. Psarris and C. A. Floudas. Improving Dynamic Operability in MIMO Systems with Time Delays. *Chemical Engineering Science*, 45:3505–3524, 1990.

L. Qui and E. J. Davison. Performance limitations of non-minimum phase systems in the servomechanism problem. *Automatica*, 29(2):337–349, 1993.

H. H. Rosenbrock. *State-space and Multivariable Theory*. Nelson, London, 1970.

L. W. Russell and J. D. Perkins. Towards a Method for Diagnosis of Controllability and Operability Problems in Chemical Plants. *Chemical Engineering Research and Design*, 65(6):453–461, 1987.

A. K. Saboo, M. Morari, and D. C. Woodcock. Design of resilient processing plants-viii. a resilience index for heat exchanger networks. *Chemical Engineering Science*, 40(8):1553–1565, 1985.

S. N. Singh. Functional Reproducibility of Multivariable Nonlinear Systems. *IEEE Transactions on Automatic Control*, AC-27(1):270–272, February 1982.

S. Skogestad and K. Havre. The use of rga and condition number as robustness measures. *Computers and Chemical Engineering*, 20:S1105–S1010, 1996.

S. Skogestad and M. Morari. Design of Resilient Processing Plants-IX. Effect of Model Uncertainty on Dynamic Resilience. *Chemical Engineering Science*, 42(7):1765–1780, 1987.

S. Skogestad and M. Morari. Effect of disturbance directions on closed-loop performance. *Industrial and Engineering Chemistry Research*, 26:2029–2035, 1987.

S. Skogestad and I. Postlethwaite. *Multivariable feedback control*. John Wiley and sons, 1996.

S. Skogestad and E. A. Wolff. Controllability measures for disturbance rejection. *Modelling, Identification and Control*, 17(3):167–182, 1996.

S. Skogestad. A Procedure for SISO Controllability Analysis - with application to design of pH processes. In *IFAC Workshop on Integration of Process Design and Control, Baltimore, Maryland, USA*, pages 23–28, June 1994.

S. Skogestad. Controllability Analysis of SISO Systems. In *IFAC Symposium, ADCHEM'94, Advanced Control of Chemical Processes, May25-27, 1994, Kyoto Research Park, Kyoto,JAPAN*, pages 471–476. International Federation of Automatic Control, May 1994.

M. Soroush. Evaluation of achievable control quality in nonlinear processes. In *IFAC Workshop on Integration of Process Design and Control, Baltimore, Maryland, USA*, pages 41–46, June 1994.

R. E. Swaney and I. E. Grossman. An index for operational flexibilty in chemical process design.1. formulation and theory. *AIChE Journal*, 31(4):621–630, 1985.

C. L. E. Swartz. A computational framework for dynamic operability assessment. In *IFAC Workshop on Integration of Process Design and Control, Baltimore, Maryland, USA*, pages 153–158, June 1994.

J. Tsinias and N. Kalouptsidis. Invertibility of Nonlinear Analytic Single-Input Systems. *IEEE Transactions on Automatic Control*, AC-28(9):931–933, September 1983.

M. Vidyasagar. Optimal Rejection of Persistent Bounded Disturbances. *IEEE Transactions on Automatic Control*, AC-31(6):527–534, June 1986.

T .T. L. Vu, P. A. Bahri, and J. A. Romagnoli. Operability considerations in chemical processes: A switchability analysis. *Computers and Chemical Engineering*, 21:S143–S148, 1997.

S. P. Walsh and T. I. Malik. Controllabilty analysis and modelling requirements: an industrial example. *Computers and Chemical Engineering*, 19S:S369–S374, 1995.

S. Walsh and J. Perkins. Integrated design of effluent treatment systems. In *Interactions between Process Design and Process Control;Ed. J. D. Perkins*, pages 107–112. Pergamon, Oxford, 1992.

S. Walsh and J. Perkins. *Advances in Chemical Engineering 23: Process Synthesis; Ed. John L. Anderson*, chapter 5: Operability and Control in Process Synthesis and Design. Academic Press, 1996.

S. P. Walsh, S. D. Chenery, P. Owen, and T. I. Malik. A case study in control structure selection for a chemical reactor. *Computers and Chemical Engineering*, 21S, 1997.

V. White, J. D. Perkins, and D. M. Espie. Switchability analysis. In *IFAC Workshop on Integration of Process Design and Control, Baltimore, Maryland, USA*, June 1994.

P. Whittle. *Optimal Control: Basics and Beyond.* John Wiley and sons, 1996.

E. A. Wolff, S. Skogestad, M. Hovd, and K. W. Mathisen. A Procedure for Controllability Analysis. In *Interactions between Process Design and Process Control, IFAC Workshop, London, UK*, pages 127–132, September 1992.

D. C. Youla, H. A. Jabr, and J. J. Bongiorno. Modern Wiener-Hopf Design of Optimal Controllers - Part II: The Multivariable Case. *IEEE Transactions on Automatic Control*, AC -21(3):319–338, June 1976.

J. C. C. Young, C. L. E. Swartz, and R. Ross. On the effects of constraints, economics and uncertain disturbances on dynamic operability assessment. *Computers and Chemical Engineering*, 20S:S677–S682, 1996.

J. G. Ziegler and N. B. Nichols. Process lags in automatic-control circuits. *Transcations of the A.S.M.E.*, 65:433–444, 1943.