Conic Sector Analysis for Digital Control Systems with Structured Uncertainty

Thesis by

Russell Lane Dailey

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1987

(Submitted November 24, 1986)

to my parents

Acknowledgements

I owe a great deal of thanks to my advisors, Dr. Peter Thompson and Dr. John Doyle, for their support and encouragement. Dr. Thompson's influence first led me into the field of control theory. He introduced me to conic sector theory and encouraged me to begin studying multirate systems. He has been instrumental in motivating me to finish this thesis. Dr. Doyle's arrival at Caltech has ushered in a world of new ideas. The analysis methods I learned from him inspired much of the content of this thesis. Illuminating conversations with him have cleared up my confusion on many occasions. Both of these men have had a great influence on my life.

I wish to thank General Motors Corporation Research Laboratories and the National Science Foundation for supporting this research, and the California Institute of Technology for its financial aid in the form of tuition grants, research assistantships, and teaching assistantships.

I also thank Cubic Corporation of San Diego for its generous financial assistance in my first year of graduate school, and for being an enjoyable place of employment for three years prior to graduate school.

A word of thanks must go to the other control theory students at Caltech, especially Ricardo Sánchez Peña and Roy Smith, for their friendship and for many interesting discussions.

I owe special gratitude to my fiancée, Pamela Derry, for her constant love and encouragement.

Abstract

This thesis presents a method which greatly reduces the conservativeness of conic sector analysis for sampled data feedback systems. The new method evaluates the stability and closed-loop performance of systems with structured uncertainty in the plant transfer function, including MIMO systems and those with multiple sampling rates. In contrast to most multirate analysis techniques, the sampling rates need not be related by rational numbers; this allows analysis when samplers are not strobed to a common clock.

The method is based on a theorem from P. M. Thompson which shows how to construct a conic sector containing a hybrid operator. Combining this theorem with the Structured Singular Value approach of J. C. Doyle, with its heavy use of diagonal scaling, provides an analysis framework for systems with multiple structured plant perturbations. Chapter 3 presents a theorem for the optimal conic sector radius in the SISO case; a MIMO extension of the the theorem completes the development of the new method. Chapter 5 gives three examples.

Chapter 6 presents a new method, based on the complex cepstrum, for synthesis of SISO rational functions to match given "target" transfer functions. The method offers complete control over stability and right half plane zeros. It solves directly for poles and zeros, avoiding the numerical sensitivity of methods which solve for polynomial coefficients. It can synthesize minimum phase functions to match a given magnitude or phase curve. In an example, it is used to synthesize a low-order digital replacement for an analog compensator which gives no degradation of

stability margin or step response.

This thesis also presents a method for Kranc vector switch decomposition in state space; this is for stability analysis and input-output simulation of perturbed multirate systems. Moving the 30-year-old Kranc technique from the frequency domain to the state-space domain simplifies the analysis tremendously. Because the number of states is preserved, the dimensionality problems long associated with the Kranc method disappear. The new method is also useful for simulating intersample ripple behavior.

Table of Contents

Acknowledgements ..... *iv*

Abstract ..... *v*

List of Illustrations ..... *xi*

1. Introduction 1

1.1 Why Use Conic Sectors? 1

1.2 Contributions of Thesis 4

1.3 Summary of Thesis 6

2. Mathematical Background 9

2.0 Table of Symbols 9

2.1 Vector Norms, Matrix Norms, and Singular Values 11

2.2 L_2 and L_{2e} Vector Spaces and MIMO Transfer Functions 13

2.2.1 State Space Interconnection Rules 16

2.3 Conic Sectors 19

2.4 Conic Sector Stability Criterion 21

2.5 The Structured Singular Value μ 28

2.6 Robustness Analysis for LTI Systems 33

2.6.1 Linear Fractional Transformations 33

2.6.2 Nominal Stability 33

2.6.3 Robust Stability 35

2.6.4 Nominal Performance 37

2.6.5 Robust Performance 38

2.7 Sampled-data Systems 41
2.7.1 Digital Filters: a Continuous-time Interpretation 41
2.7.2 z-transform Loop Stability Analysis 42
2.7.3 Notation for Sampled Signals and Systems 45
2.7.4 Hybrid Operators 46
2.7.5 A Property of PLTV Operators 48
2.8 Multi-rate Analysis: Kranc Operators in State Space 51
2.8.1 State Space Realizations for Kranc Operators 51
2.8.2 Robustness Analysis using Kranc Operators 57
3. Conic Sectors for Hybrid Operators 60
3.1 Basic Theorem: a Cone that Contains the Hybrid Operator 60
3.2 Limitations of the Basic Theorem 67
3.3 Flexibility of the Basic Theorem 69
3.3.1 A Simple Cone for a Sampler with a ZOH 72
3.4 Optimal Cone Radius in SISO Case 80
3.4.1 Theorem and Proof 83
3.4.2 Robust Stability for a System with One Sampler 89
3.5 Stability Criterion for a Cone-bounded Diagonal Operator 97
3.6 Improved Stability Criterion for a Class of Diagonal Cone-bounded Operators 99
3.7 A Nonconservative Conic Sector Method for MIMO Multirate Nonsynchronous Systems 103
3.7.1 New Method and Explanation 103
3.7.2 How Much Conservativeness is Left? 111
3.7.2.1 Systems Unstable with Cone Center 112
3.7.2.2 Optimal Radius Too Large 115
3.7.2.3 Skewed Sampling Effects 118

3.7.2.4 Sampling the Output of a ZOH 124
3.8 A Conic Sector Containing Two Samplers 131
3.9 Robustness Analysis Using New Results 137
3.9.1 Nominal Stability 137
3.9.2 Robust Stability 137
3.9.3 Nominal Performance 142
3.9.4 Robust Performance 147
4. Practical Considerations 152
4.1 Computing the Cone Center and Radius 152
4.1.1 State Space Method for SISO Operators 152
4.1.2 Numerical Method for SISO Operators 154
4.2 Matrix Balancing and Scaling 156
4.2.1 Osborne's Method 156
4.2.2 Singular Value Minimization 158
4.3 Application to Typical Control System Specifications 159
4.3.1 Gain Margin, Phase Margin, and Gain-phase Margin 159
4.3.2 Loop-at-a-time Analysis 162
4.3.3 RMS Output Error for Sinusoidal Inputs 163
4.3.4 RMS Output for White Noise Inputs 166
4.3.5 Step Response 169
5. Examples 171
5.1 SISO Single Rate Example 171
5.2 MIMO Single Rate Example 194
5.3 Multi-rate Example 205
6. Synthesis of Digital Controllers 214
6.1 The Complex Cepstrum 216
6.2 Complex Cepstrum of a Rational Function in z 217

6.3 Optimal Matching of the Complex Cepstrum 219

6.3.1 The Cost Function J and its Gradient 220

6.3.2 Real Parameterization of Poles and Zeros 221

6.3.3 Applications 223

6.4 Example: Matching an Analog Transfer Function 224

6.5 Example: Discrete-time Compensator Synthesis 229

7. Conclusion 242

7.1 Suggestions for Future Research 247

References 250

List of Illustrations

Fig. 2.4-1: A basic feedback loop 23

Fig. 2.4-2: $\bar{G}$ replaced by $\hat{G} + C$ 23

Fig. 2.4-3: $\hat{G}$ replaced by $\Delta R = \hat{G}$ 24

Fig. 2.4-4: The feedback loop containing Δ 24

Fig. 2.5-1: A feedback loop with constant M, Δ 29

Fig. 2.6-1: An upper linear fractional transformation 34

Fig. 2.6-2: A lower linear fractional transformation 34

Fig. 2.6-3: A feedback loop with disturbance inputs 36

Fig. 2.6-4: A system with "performance block" Δ_{n+1} 39

Fig. 2.6-5: Δ_{n+1} augmented with Δ 39

Fig. 2.7-1: A sampled-data feedback system 43

Fig. 2.7-2: A hybrid operator 47

Fig. 2.7-3: A hybrid operator with $D(e^{sT})$ lumped with $H(s)$ 47

Fig. 2.8-1: A multirate system 52

Fig. 2.8-2: Same system with samplers at period T 52

Fig. 2.8-3: Same system with Kranc operator 52

Fig. 2.8-4: A perturbed multirate feedback system 58

Fig. 2.8-5: Kranc operator equivalent 58

Fig. 3.1-1: A hybrid operator 61

Fig. 3.2-1: A sampled-data feedback loop 68

Fig. 3.3-1: A sampled-data feedback loop 70

Fig. 3.3-2: Insertion of α, α^{-1}, β, and β^{-1} 70

Fig. 3.3-3: A sandwiched sampler 73

Fig. 3.3-4: A sampler followed by a ZOH 73

Fig. 3.3-5: Sampler and ZOH with α and α^{-1} 75

Fig. 3.3-6: Same system with $(1 - e^{-sT})$ commuted 75

Fig. 3.3-7: Sine response of C and of sampler with ZOH 77

Fig. 3.3-8: Sampled-data loop with ZOH 79

Fig. 3.3-9: Same system with D and H_0^T commuted 79

Fig. 3.4-1: SISO system with α, α^{-1}, β, β^{-1} 81

Fig. 3.4-2: K split into LTI part and PLTV part 81

Fig. 3.4-3: LTI portion lumped into M 82

Fig. 3.4-4: The simplified system 82

Fig. 3.4-5: SISO system with model uncertainty 90

Fig. 3.4-6: System after inserting α and β 90

Fig. 3.4-7: Extraction of $\hat{\Delta}$ inside Cone$(0, I)$ 92

Fig. 3.4-8: Δ and $\hat{\Delta}$ lumped into a diagonal operator 92

Fig. 3.4-9: The composite operator $\hat{P}$ 95

Fig. 3.5-1: Feedback loop with cone-bounded diagonal K 98

Fig. 3.6-1: System with D, D^{-1} inserted 100

Fig. 3.6-2: Simplified system 100

Fig. 3.7-1: A MIMO sampled-data loop 104

Fig. 3.7-2: System after inserting A, A^{-1}, B, and B^{-1} 104

Fig. 3.7-3: Insertion of $R^{1/2}$ and $R^{-1/2}$ 107

Fig. 3.7-4: $R^{1/2}$ commutes with K 107

Fig. 3.7-5: A stable system, unstable with cone center 113

Fig. 3.7-6: Log Nyquist plots of $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$ 113

Fig. 3.7-7: Nyquist plots of $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$ 116

Fig. 3.7-8: Optimal radius R 116

Fig. 3.7-9: Bode plot of $\tilde{M} = GH_0^T(1 - \frac{1}{T}GH_0^T)^{-1}$ 117

Fig. 3.7-10: System stable with skewed sampling 117

Fig. 3.7-11: Nyquist plots of three loop gains 120

Fig. 3.7-12: MIMO Nyquist test: $\det(I - \hat{T}^{-1}M)$ 120

Fig. 3.7-13: Bode plot of $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ 122

Fig. 3.7-14: An analog feedback loop 125

Fig. 3.7-15: Step response for analog and multirate systems 125

Fig. 3.7-16: Multirate version of Fig. 3.7-14 127

Fig. 3.7-17: Log Nyquist plot of $\det(I - \hat{T}^{-1}M)$ 127

Fig. 3.7-18: Bode plots of α_1 and β_1 129

Fig. 3.8-1: A SISO multirate system 132

Fig. 3.9-1: MIMO system with structured plant uncertainty 138

Fig. 3.9-2: Construction of $\tilde{M}$ for robust stability test 138

Fig. 3.9-3: Construction of diagonal K inside $\mathrm{Cone}(0, I)$ 140

Fig. 3.9-4: Feedback loop with two relations $\mathcal{K}$ and $\mathcal{G}$ 144

Fig. 3.9-5: MIMO system with external inputs and outputs 146

Fig. 3.9-6: Construction of $\tilde{M}$ for nominal performance test 146

Fig. 3.9-7: Uncertain MIMO system with external inputs and outputs 148

Fig. 3.9-8: Construction of $\tilde{M}$ for robust performance test 150

Fig. 4.3-1: Calculation of gain and phase margins 160

Fig. 4.3-2: Calculation of gain-phase margin 160

Fig. 5.1-1: A SISO single-rate system 172

Fig. 5.1-2: The rearranged system 172

Fig. 5.1-3: System after cone bounding 174

Fig. 5.1-4: Bode plot of radius $R = \mathrm{Rad}(\alpha, \beta)$ 174

Fig. 5.1-5: Log Nyquist plots of $\frac{1}{T}FGH_0^TK$ and $K(FGH_0^T)^T$ 175

Fig. 5.1-6: Construction of $\hat{M}$ for nominal performance test 177

Fig. 5.1-7: Bode plots of C and R_2 179

Fig. 5.1-8: System with multiplicative uncertainty 179

Fig. 5.1-9: System restructured for robust stability test 181

Fig. 5.1-10: Construction of $\hat{M}$ for robust stability test 181

Fig. 5.1-11: Bode plots of $|1 + L^{-1}(z)|$ and $W_{\max}$ 183

Fig. 5.1-12: System restructured for robust performance test 183

Fig. 5.1-13: Construction of $\hat{M}$ for robust performance test 185

Fig. 5.1-14: Bode plots of C and R_2 187

Fig. 5.1-15: Behavior of α in gain margin test 187

Fig. 5.1-16: Behavior of α in phase margin test 188

Fig. 5.1-17: System restructured for exact RMS calculation 188

Fig. 5.1-18: Bode plots of $|C| + |R|$ and actual RMS gain 191

Fig. 5.1-19: Actual and estimated step response 191

Fig. 5.2-1: MIMO system with multiplicative uncertainty 195

Fig. 5.2-2: State space quadruples for $G(s)$ and $K(z)$ 196

Fig. 5.2-3: Equivalent system with constant ΔW 198

Fig. 5.2-4: MIMO log Nyquist plot: $\det(I + (FGH_0^T)^T K)$ 198

Fig. 5.2-5: System rearranged for conic sector analysis 200

Fig. 5.2-6: Construction of $\hat{M}$ for robust stability test 200

Fig. 5.2-7: Bode plot of $\bar{\sigma}(R)$ 202

Fig. 5.2-8: Robust stability margins 202

Fig. 5.2-9: Alternative robust stability margin 204

Fig. 5.3-1: A SISO multirate system 204

Fig. 5.3-2: Robust stability margins 207

Fig. 5.3-3: System rearranged for conic sector analysis 207

Fig. 5.3-4: Bode plots of α_1 and α_2 210

Fig. 5.3-5: SISO cone radii R_1 and R_2 210

Fig. 5.3-6: Bode plot of $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}_{\rm ul}B^{-1}R^{1/2})$ 211

Fig. 5.3-7: Log Nyquist plot of $L = \frac{6}{T^2}FKH_0^{T/2}FGH_0^{T/3}$ 211

Fig. 5.3-8: Construction of $\hat{M}$ for robust stability test 212

Fig. 6.1-1: A SISO hybrid controller 215

Fig. 6.4-1: Bode plots of C and $\hat{C}$ 225

Fig. 6.4-2: Log transfer function $\hat{X}(e^{j\omega'})$ 227

Fig. 6.4-3: The complex cepstrum $\hat{x}(n)$ 228

Fig. 6.4-4: Bode plots of target C and error $C - P$ 228

Fig. 6.5-1: An analog feedback system 231

Fig. 6.5-2: System with $K_a(s)$ replaced by hybrid controller 231

Fig. 6.5-3: Bode plots of target function 232

Fig. 6.5-4: Complex cepstrum $\hat{x}(n)$ 235

Fig. 6.5-5: Bode plots of target and causal best fit 236

Fig. 6.5-6: Bode plots of best fits to phase-corrected target 238

Fig. 6.5-7: Bode plots of loop gains K_aG and $\frac{1}{T}H_0^TK_dFG$ 239

Fig. 6.5-8: Step responses of analog and hybrid systems 241

Chapter One — Introduction

1.1 Why use conic sectors?

Since the invention of the microprocessor, control system designers have put the device to use on a wide scale in computer-controlled feedback systems. Digital controllers have many advantages over their analog counterparts. They are easily reprogrammed without expensive wiring changes; they are immune to parameter drift over temperature or time; they are smaller in size and weight; they are often much cheaper than analog controllers.

Because of the time-varying nature of the sampling operation of an A-D converter, however, a digitally controlled feedback system is not linear time-invariant (LTI). A transfer function cannot be defined, since the impulse response varies depending on the position of the impulse within a sampling time interval. Therefore, conventional analysis methods based on the Laplace transform cannot be used.

To circumvent this problem, analysis methods based on the z-transform were developed. The analog portions of a system are considered only at the outputs of samplers; in this way they can be treated as linear shift-invariant discrete-time operators, for which transfer functions are defined. Analysis of stability and time response proceeds easily once a transfer function for the discretized analog plant is found.

One drawback of the z-transform approach lies in evaluating safety margins for *robust* stability; that is, determining how much the analog plant can vary before the

system becomes unstable. In practice, the plant transfer function is never known exactly; a model is chosen to approximate the plant. The modelling error is usually small at low frequencies and quite large at high frequencies. When the analog plant model is discretized, however, the result is valid only if the plant exactly equals the model at all frequencies. Attempts to carry frequency-dependent estimates of modelling error over into the discrete-time domain are hampered by frequency aliasing in the discretization process.

Another drawback appears when conventional z-transform methods are attempted for *multi-rate* systems: systems with multiple samplers operating at different speeds. Since the various discrete-time portions of such systems run in different time frames, they are not shift invariant with respect to each other; no transfer function exists to relate one to the other. The only way out of this dilemma which remains within the z-transform framework is to find a common time frame for all samplers. This is possible if the sampling rates are related by rational numbers; for instance, if one sampler has period $T/2$ and another has period $T/3$, they can both be modelled by samplers of period T with appropriate time delays and advances. This is the objective of Kranc vector switch decomposition [12],[2]. Other methods have been developed for the special case of sampling rates related by integers, but these are mathematically equivalent to the Kranc method. Because Kranc switch decomposition was originally formulated in the frequency domain (in terms of rational functions in z), it was awkward to use and generally avoided. Section 2.8 of this thesis repeats results given earlier in [22], [21] which show that, when recast into a state-space format, the Kranc method becomes simple and easy to use.

If the sampling rates are not related by rational numbers, however, the Kranc method and other z-transform methods break down completely. This can happen in practice when samplers are free-running and not strobed to a common clock. Furthermore, the Kranc method has the same weakness as other z-transform techniques

when robust stability analysis is attempted.

One more drawback of z-transform methods lies in evaluating the continuous-time response of a system to continuous-time inputs, e.g., the closed-loop RMS output to a sine input. Since z-transform techniques view the analog plant only at the output of the samplers, such continuous-time information is lost.

All of these drawbacks are consequences of discretization. All can be avoided by switching to an analysis method which treats discrete-time compensators as *continuous-time* operators. By remaining in the continuous time domain, we can use frequency-dependent modelling error information with no conservativeness from discretization; we can analyze multi-rate systems, even when the sampling rates are related by irrational numbers, and we can model the input-output behavior of systems from analog inputs to analog outputs. The price of using continuous time analysis is that the system is not LTI; it has no transfer function. However, *conic sector* methods allow us to model such a system with an LTI transfer function approximation called the *cone center*, and to account strictly for modelling errors with another LTI transfer function called the *cone radius*. Since the cone radius accounts strictly for the modelling error, we can make definitive statements about system stability and performance. Conic sectors let us perform rigorous analysis of non-LTI systems using only two LTI transfer functions.

Conic sector analysis was originated by Zames in 1966 [27], [28]. In Zames's original formulation, the cone center and radius both had to be constant multipliers; frequency dependent information was lost. Later, Safonov [16] greatly generalized Zames's results to apply when the cone center and radius are dynamic operators, such as transfer functions.

Conic sector methods were first applied to sampled data operators, or hybrid operators, by Kostovetsky [11], and later by Stein [19]. In both these results the cone radius was a constant multiplier, which led to conservativeness due to its lack

of frequency-dependent information. Thompson [20],[29] then discovered how to compute a frequency-dependent cone center *and* radius for any open-loop stable, single-rate hybrid compensator. This thesis extends the results of [20], removing several restrictions and reducing conservativeness, especially for multivariable systems.

1.2 Contributions of Thesis.

There are three chief contributions of this thesis:

Improved conic sector theory for sampled-data systems. The improvements in the theory presented here are all based on Theorem 3.2 of [20], which shows how to compute the cone center and radius for a hybrid operator consisting of a prefilter, sampler, digital filter, and a hold element. It was realized early by Enns and Doyle [7],[4] that this theorem allowed for a great deal of flexibility, since fictitious LTI operators could be substituted for the prefilter and hold. This allowed the cone radius to be manipulated at will, and the conservativeness of the conic sector loop stability test to be reduced. This thesis presents an optimal choice for these fictitious operators in the SISO case; that is, a choice which minimizes the conservativeness of the nominal stability test.

For the MIMO case, the optimal SISO result is extended in a way which, though suboptimal, gives nonconservative results for a wide class of systems. The theory applies equally to single-rate and multi-rate MIMO systems, even if the samping rates are not related by rational numbers. The MIMO approach used here relies heavily on diagonal scaling of matrices and is based on the Structured Singular Value (SSV) analysis methods of Doyle [5],[6]. Like the SSV techniques, this method allows for robustness analysis of systems with structured plant uncertainty; i.e., systems with several different model perturbations, all of which are allowed to vary

independently.

The new theory not only tests for nominal and robust system stability, but also tests for closed-loop performance under nominal and perturbed conditions. It does this by constructing a conic sector containing the closed-loop system. This conic sector can be used to determine if the system meets specifications on closed-loop response to sine waves, white noise, and other inputs.

Complex cepstrum synthesis method for sampled-data SISO compensators. This is a technique for converting analog compensators into hybrid compensators which will provide the same closed-loop response. The philosophy is the same as in previous work by Thompson, Sánchez Peña, and Wong [24]: to synthesize a digital filter $K_d(z)$ so that the hybrid compensator's cone center $\frac{1}{T}H(j\omega)K_d(e^{j\omega T})F(j\omega)$ approximates the analog compensatior $K_a(s)$ at frequencies inside the loop bandwidth. H and F are the hold and prefilter, respectively, and T is the sampling period. The approach taken here is to construct a target function defined on the unit circle, compute its complex cepstrum, and then use that to synthesize a rational function in z of specified order which approximates it.

The complex cepstrum function was developed in the digital signal processing community, and has been used previously in applications such as pole-zero modelling of the human vocal tract and echo removal from old phonograph recordings [14]. It has the very desirable property of separating the minimum phase and maximum phase portions of a transfer function: minimum phase functions are mapped into causal time sequences, and maximum phase functions into anticausal time sequences. It also has a simple parameterization in terms of the poles and zeros of a rational function. These two properties together give us complete control over the number of poles and zeros inside and outside the unit circle in the synthesis procedure. We can also solve directly for pole and zero locations, rather than for the coefficients of the expanded numerator and denominator polynomials. This avoids

problems with the notorious sensitivity of high-order polynomials to tiny changes in their coefficients.

This complex cepstrum synthesis procedure has other applications, including fitting rational functions in s or z to measured frequency response data, generating phase curves from magnitude information only, and generating magnitude curves from phase curves. The relationship between magnitude and phase holds only for minimum phase functions.

Multi-rate analysis using Kranc operators in state space. This thesis presents a minimal state-space realization for the MIMO linear shift invariant operators which appear when the Kranc vector switch decomposition technique [12],[2],[26] is used. The Kranc method applies generally to all multi-rate sampled-data systems which have sampling rates related by rational numbers. It has been avoided for thirty years because, when performed in the frequency domain using rational functions in z, it is awkward and difficult. When recast into a state-space format, however, it becomes transparently simple and easy to perform with appropriate software. The only "sophisticated" numerical operations needed to construct the state-space quadruples are matrix exponentiation and matrix multiplication. Testing for stability involves only computing the eigenvalues of the closed-loop state transition matrix. Input-output time domain simulation is trivially easy with a state-space representation.

This material has appeared earlier in [22]. Thompson's method for computing gain and phase margins using Kranc operators is also presented; this appeared earlier in [20].

1.3 Summary of Thesis.

Chapter 1. Introduction.

Chapter 2. Mathematical background. This covers L_2 and L_{2e} vector spaces, matrix norms and singular values, conic sector definitions, the conic sector stability criterion, standard analysis techniques for sampled data systems, and the new material on Kranc operators in state space. Also covered is the Structured Singular Value μ and its applications for testing nominal stability, robust stability, nominal performance, and robust performance for analog LTI feedback systems.

Chapter 3. Conic sectors for hybrid operators. This presents the new conic sector theory. It begins by repeating Theorem 3.2 from [20] and its proof, upon which all the new theory is based. It discusses the limitations of the theorem and the flexibility inherent in it. A simple conic sector containing only a sampler and a zero order hold (ZOH) is described; this has appeared previously in [23]. Then the optimal radius for testing nominal stability of a SISO hybrid system is given, together with a proof of its optimality. The conceptual groundwork for treating MIMO systems with diagonal scaling is laid, then the MIMO result is described: a (usually) nonconservative conic sector method for multi-sampler MIMO systems, including multi-rate systems. Considerable discussion and several examples are then devoted to examining the limits of the new technique: finding cases in which the new method still gives conservative results. It appears that conservativeness usually appears only when a system has poor anti-aliasing filtering, poor robustness margins, or both. An exception occurs with multi-rate systems having a sampler at the output of a ZOH: a modified conic sector theorem is derived for this case.

Finally, Section 3.9 describes how to perform robustness analysis tests using the new results. The tests are grouped in the same way as those for SSV analysis [6]: nominal stability, robust stability, nominal performance, and robust performance.

Chapter 4. Practical considerations. This chapter begins by describing two ways to compute the conic sector radius using Theorem 3.2 of [20]. Osborne's matrix balancing method and its application to computing the SSV and the MIMO

conic sector radius are discussed. Section 4.3 describes how to convert typical control system specifications into a form compatible with the conic sector methods of this thesis. These include gain margin, phase margin, gain-phase margin, RMS output for sine and white noise inputs, and step response. All of these except step response are directly compatible with conic sector methods. Even in this last case, it is shown by example that the cone center can provide a good approximation to the closed loop step response.

Chapter 5. Examples. Three examples are worked through in detail: a SISO system, a MIMO single-rate system, and a multi-rate system. The first two examples were used previously in [20] with the earlier version of the conic sector theory. The first example is worked in great detail: nearly every test described in this thesis is performed on it. For the last two examples, nominal stability and a robust stability margin are tested. The first two examples prove to give extremely nonconservative results. The last example is slightly conservative, but still gives a more useful robust stability margin than Kranc techniques.

Chapter 6. Complex cepstrum synthesis method. This new synthesis method for hybrid controllers is described. The chapter begins by defining the complex cepstrum and describing its properties. Then a nonlinear optimization problem to solve for poles and zeros is set up; the cost function and gradient are shown to be simple functions of the pole and zero locations. A real parameterization is introduced for poles and zeros which speeds the computation and guarantees that complex poles and zeros always appear in complex conjugate pairs. Finally, an example of the procedure is given; the result is a low-order hybrid compensator which gives an excellent match to the target analog compensator. Another example illustrates how to use this technique to fit rational functions to analog frequency response data.

Chapter 7. Conclusion.

Chapter Two — Mathematical Background

2.0 Table of Symbols

R	the real numbers
R^n	real Euclidean space of dimension n
$R^{n\times m}$	the set of $n \times m$ matrices with real elements
C	the complex numbers
C_+	the complex right half-plane
C_-	the complex left half-plane
C^n	complex Euclidean space of dimension n
$C^{n\times m}$	the set of $n \times m$ matrices with complex elements
A^T, x^T	transpose of matrix A or vector x
A^*, x^*	Hermitian transpose of matrix A or vector x
trace(A)	the sum of the diagonal elements of matrix A
$\lambda_i(A)$	the ith eigenvalue of matrix A
$\sigma_i(A)$	the ith singular value of matrix A
$\bar{\sigma}(A)$	maximum singular value of matrix A
$\underline{\sigma}(A)$	minimum singular value of matrix A
$\rho(A)$	spectral radius of matrix A
$\|x\|$	Euclidean norm of vector x in R^n or C^n
$\|A\|_2$	induced Euclidean norm of matrix A
$\|A\|_F$	Frobenius norm of matrix A
L_2	square-integrable vector-valued functions of a real variable

L_{2e}	extended L_2 space
$\langle x, y \rangle$	inner product in L_2: $\int_0^\infty x^*(t)y(t)dt$
$\langle x, y \rangle_\tau$	truncated inner product: $\int_0^\tau x^*(t)y(t)dt$
P_τ	truncation operator
$\|x\|$	$\langle x, x \rangle^{1/2}$ for $x \in L_2$
$\|x\|_\tau$	$\langle x, x \rangle_\tau^{1/2}$ for $x \in L_{2e}$
H_2	functions analytic in C_+ and square-integrable on $j\omega$-axis
$H_2^\perp$	functions analytic in C_- and square-integrable on $j\omega$-axis
H_∞	functions analytic in C_+ and bounded on $j\omega$-axis
L_∞	functions bounded on $j\omega$-axis
$\|M\|_\infty$	$\sup_{\omega \in R} \bar{\sigma}(M(j\omega))$
LTI	linear time-invariant
LTV	linear time-varying
PLTV	linear and periodically time varying
SISO	single input single output
MIMO	multiple input multiple output
$\Leftrightarrow$	if and only if
$\exists$	there exists
$\forall$	for all
$F_l(J, Q)$	$J_{11} + J_{12}Q(I - J_{22}Q)^{-1}J_{21}$ (lower fractional transformation)
$F_u(J, Q)$	$J_{22} + J_{21}Q(I - J_{11}Q)^{-1}J_{12}$ (upper fractional transformation)
$(G)^T(z)$	sample equivalence operator for $G(s)$
H_0^T	$(1 - e^{-sT})/s$
ZOH	a zero-order hold H_0^T

2.1 Vector Norms, Matrix Norms and Singular Values

For a vector x in C^n or R^n, the *Euclidean norm* is defined by

$$\|x\| \;=\; \|x\|_2 \;=\; \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

where x_i are the components of x. The symbol $\|x\|$, where x is a vector, denotes the Euclidean norm of x throughout this thesis. For a matrix $A = [a_{ij}]$ in $C^{m\times n}$ or $R^{m\times n}$, two norms are used in this thesis. The *Frobenius norm* $\|A\|_F$ is defined by

$$\|A\|_F \;=\; \sqrt{\text{trace}(A^*A)} \;=\; \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}$$

The Frobenius norm is often used for the practical reason that it is easy to compute, since $\|A\|_F^2$ is simply the sum of the squared magnitudes of all the elements of A.

The *induced* Euclidean matrix norm, or 2-norm, is given by

$$\|A\|_2 \;=\; \max_{x\neq 0} \frac{\|Ax\|}{\|x\|}$$

where x ranges over all $x \in C^n$. The 2-norm may also be defined in terms of the *singular values* of a matrix. The singular values of a matrix A in $C^{m\times n}$ or $R^{m\times n}$ are given by

$$\sigma_i(A) \;=\; \sqrt{\lambda_i(A^*A)}$$

where the λ_i are the eigenvalues of A^*A for $i = 1$ to $\min(m,n)$, and the nonnegative square root is taken. By convention, the singular values are numbered in descending order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. It can be shown that the largest singular value is equal to the 2-norm of the matrix:

$$\bar{\sigma}(A) \;=\; \|A\|_2$$

where $\bar{\sigma}$ denotes the largest singular value. $\underline{\sigma}$ is defined similarly as the smallest singular value. Since $\bar{\sigma}(A)$ is the highest possible ratio between output norm $\|Ax\|$

and input norm $\|x\|$, it is also called the *gain* of the matrix A. $\|A\|_F$ and $\|A\|_2$ are compatible norms; it can be shown that for any $n \times n$ matrix A,

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$$

The following inequality is sometimes useful for square matrices A:

$$\underline{\sigma}(A) \leq \min_i |\lambda_i(A)| \leq \max_i |\lambda_i(A)| \leq \bar{\sigma}(A)$$

In other words, $\underline{\sigma}(A)$ and $\bar{\sigma}(A)$ are lower and upper bounds, respectively, for the magnitudes of the smallest and largest eigenvalues of A. $\max_i |\lambda_i(A)|$ is also called $\rho(A)$, the *spectral radius* of A.

Any $m \times n$ matrix A has a singular value decomposition (SVD) given by

$$A = U\Sigma V^*$$

where Σ is an $m \times n$ matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and U and V are unitary matrices of dimension $m \times m$ and $n \times n$, respectively. The SVD partitions the domain and range of A into orthogonal subspaces which correspond to the orthonormal columns of U and V. If u_i and v_i are the ith columns of U and V, respectively, then

$$Av_i = U\Sigma V^* v_i = U\Sigma \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = U \begin{pmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{pmatrix} = \sigma_i u_i$$

So the SVD assigns, to each orthogonal input "direction" v_i, a gain of σ_i and a corresponding output "direction" u_i. If $n > m$ (more columns than rows), then $Av_i = 0$ for $i > m$, and the extra columns of V form an orthonormal basis for the

nullspace, or kernel, of A. If $m > n$, the extra columns of U (u_i for $i > n$) all lie in spatial directions which cannot be reached by any input to A. Therefore, these columns of U form an orthonormal basis for the orthogonal complement of the range of A; i.e. they are orthogonal to all possible values of the matrix "output" Ax.

Singular values, and the entire SVD, are readily computed by reliable, numerically well-conditioned computer routines. Because of their many desirable properties, they are frequently used in present-day control theory.

2.2 L_2 and L_{2e} Vector Spaces and MIMO Transfer Functions.

The Lebesgue space L_2 is used extensively in the conic sector theory which follows. L_2 is the space of square-integrable, causal vector-valued functions of time. Stated formally:

$$L_2 \equiv \{x(t) \mid x(t) \in C^n,\ t \geq 0, \text{ and } \|x\| \text{ is finite}\}$$

where for $x \in L_2$,

$$\|x\| \equiv \sqrt{\int_0^\infty x^*(t)x(t)dt} \equiv \langle x, x\rangle^{1/2}$$

The squared norm $\|x\|^2$ is the squared Euclidean norm of the vector $x(t)$, integrated over time. $\|x\|^2$ is sometimes called the *energy* of the signal x. Note that this definition implies that $x(t) \to 0$ as $t \to \infty$ for all $x \in L_2$. However, many signals we are interested in do not satisfy this condition; for example, the impulse response of an integrator. To handle such cases, it is necessary to introduce the extended vector space L_{2e}. First, a few definitions are needed. Let P_τ be the *linear truncation operator* defined as follows:

$$(P_\tau x)(t) = \begin{cases} x(t) & \text{if } t \leq \tau \\ 0 & \text{if } t > \tau \end{cases}$$

In other words, P_τ simply truncates a time function at $t = \tau$. Now define

$$\langle x, y\rangle_\tau \equiv \langle P_\tau x, P_\tau y\rangle = \int_0^\tau x^*(t)y(t)dt$$

$$\|x\|_\tau \equiv \langle x, x\rangle_\tau^{1/2}$$

Now we can define the extended space L_{2e}:

$$L_{2e} \equiv \{x \mid P_\tau x \in L_2 \quad \forall\, \tau \geq 0\}$$

L_{2e} is therefore the set of vector-valued functions of time which do not "blow up" at any *finite* value of time; but it includes unbounded functions such as $x(t) = t^2$ which are finite for all finite values of t. L_{2e} includes the impulse responses of all unstable, strictly proper rational transfer functions; e.g., $1/(s-1)$, which has impulse response e^t.

A causal, linear time-invariant (LTI) analog system, or convolution operator, has the following input-output behavior in the time domain:

$$y(t) = h(t) * u(t) \equiv \int_{-\infty}^{\tau} h(t-\tau)u(\tau)d\tau$$

where u is the input, y is the output, and h is the impulse response. The definition is valid for multiple-input multiple-output (MIMO) systems when $u(t)$ and $y(t)$ are vector-valued and $h(t)$ is a matrix-valued function of time.

The transfer function is given by

$$H(j\omega) = \int_{-\infty}^{\infty} e^{-j\omega t}h(t)dt$$

In the MIMO case, each element of $H(j\omega)$ is the Fourier transform of the corresponding element of $h(t)$. In the frequency domain, the input-output behavior is

$$Y(j\omega) = H(j\omega)U(j\omega)$$

where $Y(j\omega)$ and $U(j\omega)$ are the Fourier transforms of the output and input. Usually the Laplace transform notation is used: $Y(s) = H(s)U(s)$.

Most of the analog transfer functions used in this thesis can be described in state space:

$$\frac{d}{dt}x = Ax + Bu$$

$$y = Cx + Du$$

where A, B, C, and D are real matrices. It is easily shown that the transfer function of such a system is given by

$$H(s) = C(sI - A)^{-1}B + D$$

We will use the following convenient notation for transfer functions of this type:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \equiv C(sI - A)^{-1}B + D$$

This state-space representation is especially useful when computing the transfer function of an *interconnection* of several systems. Section 2.2.1 below lists several useful rules, using this notation, for system interconnection and other commonly performed block diagram manipulations.

The *poles* of this type of transfer function are defined as those complex values of s for which $\det(sI - A) = 0$. If the transfer function $H(s)$ is single-input single-output (SISO), its *zeros* are defined as the complex values of s for which $H(s) = C(sI - A)^{-1}B + D = 0$. For MIMO systems, there are many definitions for zeros, each with its own proponents and detractors. We will not be concerned with MIMO zeros in this thesis.

The following norm will be used extensively in the following pages:

$$\|H\|_\infty \equiv \sup_{\omega \in R} \bar{\sigma}(H(j\omega))$$

The ∞-norm of a MIMO transfer function is the maximum gain, over all frequencies, of the transfer function matrix. This has many interpretations. One is that $\|H\|_\infty$ is the maximum gain in (Euclidean) amplitude for a sine-wave input: if the input is

$u(t) = u_0 e^{j\omega t}$ where u_0 is a constant vector, then the output is $y(t) = (H(j\omega)u_0)e^{j\omega t}$. The maximum gain is then

$$\sup_{\omega \in R, u_0 \in C^n} \frac{\|H(j\omega)u_0\|}{\|u_0\|} \;=\; \sup_{\omega \in R} \bar{\sigma}(H(j\omega)) \;=\; \|H\|_\infty$$

$\|H\|_\infty$ is also the maximum gain in the L_2-norm for the set of all inputs in L_2; that is, for all inputs of finite energy. This can be proved as follows: let $y = Hu$, with $u \in L_2$. Since H is LTI we can find y in terms of the Fourier transform of u: $Y(j\omega) = H(j\omega)U(j\omega)$. Now

$$\|y\|^2 \;=\; \int_0^\infty y^*(t)y(t)dt \;=\; \frac{1}{2\pi}\int_{-\infty}^\infty Y^*(j\omega)Y(j\omega)d\omega$$

by Parseval's theorem. Since for fixed ω,

$$Y^*(j\omega)Y(j\omega) \;=\; \|Y(j\omega)\|^2 \;\leq\; \bar{\sigma}^2(H(j\omega))\|U(j\omega)\|^2$$

we get

$$\begin{aligned}\|y\|^2 \;&=\; \frac{1}{2\pi}\int_{-\infty}^\infty \|Y(j\omega)\|^2 d\omega \;\leq\; \frac{1}{2\pi}\int_{-\infty}^\infty \bar{\sigma}^2(H(j\omega))\|U(j\omega)\|^2 d\omega \\ &\leq\; \frac{1}{2\pi}\int_{-\infty}^\infty \|H\|_\infty^2 U^*(j\omega)U(j\omega)d\omega \;=\; \|H\|_\infty^2\|u\|^2\end{aligned}$$

This shows that $\|H\|_\infty \geq \frac{\|Hu\|}{\|u\|}$ for all nonzero $u \in L_2$. This property of the ∞-norm will be used in Section 2.4.

2.2.1 State space interconnection rules.

This section presents the state-space operations for several common block diagram manipulations. This material has appeared earlier in the Appendix of [22]. The multi-input multi-output descriptions of $G_i(s)$ or $G_i(z)$ are represented by the state-space quadruple:

$$G_i \;=\; \left[\begin{array}{c|c} A_i & B_i \\ \hline C_i & D_i \end{array}\right]$$

Multiplication:

$$G_1 G_2 = \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right] \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right]$$

$$= \left[\begin{array}{cc|c} A_1 & B_1 C_2 & B_1 D_2 \\ 0 & A_2 & B_2 \\ \hline C_1 & D_1 C_2 & D_1 D_2 \end{array}\right] = \left[\begin{array}{cc|c} A_2 & 0 & B_2 \\ B_1 C_2 & A_1 & B_1 D_2 \\ \hline D_1 C_2 & C_1 & D_1 D_2 \end{array}\right]$$

Note: this realization may not be minimal.

Addition:

$$G_1 + G_2 = \left[\begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & D_1 + D_2 \end{array}\right]$$

Subtraction:

$$G_1 - G_2 = \left[\begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & -B_2 \\ \hline C_1 & C_2 & D_1 - D_2 \end{array}\right]$$

Scalar Multiplication:

$$\alpha G = \left[\begin{array}{c|c} A & \alpha B \\ \hline C & \alpha D \end{array}\right]$$

Inversion:

If D^{-1} exists then

$$G^{-1} = \left[\begin{array}{c|c} A - BD^{-1}C & -BD^{-1} \\ \hline D^{-1}C & D^{-1} \end{array}\right]$$

Feedback Interconnections:

$$(I+G)^{-1} = \left[\begin{array}{c|c} A-B(I+D)^{-1}C & -B(I+D)^{-1} \\ \hline (I+D)^{-1}C & (I+D)^{-1} \end{array}\right]$$

$$G(I+G)^{-1} = \left[\begin{array}{c|c} A-B(I+D)^{-1}C & B(I+D)^{-1} \\ \hline (I+D)^{-1}C & (I+D)^{-1}D \end{array}\right]$$

$$G_1(I+G_2G_1)^{-1} = \left[\begin{array}{cc|c} A_1-B_1ED_2C_1 & -B_1EC_2 & B_1E \\ B_2EC_1 & A_2-B_2ED_1C_2 & B_2ED_1 \\ \hline EC_1 & -ED_1C_2 & ED_1 \end{array}\right]$$

where $E = (I+D_1D_2)^{-1}$

Paralleling:

$$\left[\begin{array}{cc} G_1 & 0 \\ 0 & G_2 \end{array}\right] = \left[\begin{array}{cc|cc} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ \hline C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{array}\right]$$

Adjoint:

$$G^T(-s) = \left[\begin{array}{c|c} -A^T & C^T \\ \hline -B^T & D^T \end{array}\right]$$

2.3 Conic Sectors

To define conic sectors, we need the concepts of a *relation* and its *inverse*. In general, a relation $\bar{G}$ is any set of the form $\bar{G} \subset \mathcal{X} \times \mathcal{Y}$; that is, any subset of the Cartesian product of any two sets $\mathcal{X}$ and $\mathcal{Y}$ [16]. For the purposes of this thesis, we will usually take $\mathcal{X}$ and $\mathcal{Y}$ to be L_{2e}. A relation can be thought of as the complete set of input-output pairs (x, y) for a system. An *operator* is a relation with only one output y defined for each input x.

The inverse of a relation $\bar{G} \subset \mathcal{X} \times \mathcal{Y}$ is denoted by $\bar{G}^I \subset \mathcal{Y} \times \mathcal{X}$ and is defined:

$$\bar{G}^I \equiv \{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid (x, y) \in \bar{G}\}$$

In other words, if $\bar{G}$ is the set of input-output pairs of a system, $\bar{G}^I$ is simply that same set with the order reversed. Note that the inverse of a relation *always* exists, even when a system has the same output for more than one input (noninvertible, in the usual sense), or when it has more than one output for the same input. The inverse of an operator, however, does not always exist. If the operator has the same output for more than one input, the inverse cannot be defined. A singular matrix is an example of a noninvertible operator.

A *conic sector* is defined as follows:

$$\text{Cone}(C, R) \equiv \{(x, y) \in L_{2e} \times L_{2e} \mid \|y - Cx\|_\tau \leq \|Rx\|_\tau \quad \forall\, \tau \geq 0\}$$

where C and R are linear time-invariant (LTI) operators. C is called the *cone center* and R is called the *cone radius*. A relation $\bar{G}$ can be said to be inside, strictly inside, outside, or strictly outside a given conic sector:

(a) A relation $\bar{G}$ is said to be *inside* $\text{Cone}(C, R)$ if $\bar{G} \subset \text{Cone}(C, R)$; that is, if

$$\|y - Cx\|_\tau \leq \|Rx\|_\tau \quad \forall\, \tau > 0, \ \forall\, (x, y) \in \bar{G}.$$

(b) $\bar{G}$ is *strictly inside* $\text{Cone}(C, R)$ if $\exists\, \epsilon > 0$ such that

$$\|y - Cx\|_\tau^2 \leq \|Rx\|_\tau^2 - \epsilon(\|x\|_\tau^2 + \|y\|_\tau^2) \quad \forall\, \tau \geq 0, \quad \forall\, (x, y) \in \bar{G}$$

(c) We say $\bar{G}$ is *outside* Cone(C, R) if

$$\|y - Cx\|_\tau \geq \|Rx\|_\tau \quad \forall\, \tau \geq 0,\ \forall\, (x, y) \in \bar{G}$$

(d) Finally, $\bar{G}$ is said to be *strictly outside* Cone(C, R) if $\exists\ \epsilon > 0$ such that

$$\|y - Cx\|_\tau^2 \geq \|Rx\|_\tau^2 + \epsilon(\|x\|_\tau^2 + \|y\|_\tau^2) \quad \forall\, \tau \geq 0, \quad \forall\, (x, y) \in \bar{G}$$

Observe that a relation $\bar{G}$ need not be either inside or outside any given Cone(C, R); it could satisfy $\|y - Cx\|_\tau^2 \geq \|Rx\|_\tau^2$ for some $(x, y) \in \bar{G}$ and $\tau \geq 0$, and fail to satisfy it for some other (x, y) and τ.

A conic sector can be considered a rigorous linear approximation to a system's input-output behavior. The system itself may be nonlinear, may have a time-varying impulse response, or may have uncertain parameters such as plant variations over temperature. If the system's response is inside a conic sector Cone(C, R), the cone center C provides an approximation of the system's output for any input. The cone radius R makes the approximation rigorous by providing an upper bound on how incorrect this approximation may be, in the form of a bound on the norm (the energy) of the difference $y - Cx$ between the predicted output Cx and the actual output y. In this way the complicated description of the system's input-output behavior is replaced by two LTI transfer functions.

Since analysis techniques for LTI systems are very well developed, and since LTI systems have desirable properties such as the superposition principle and the existence of transfer functions, it can simplify an analysis considerably to replace nonlinear and time-varying elements with conic sectors. The price paid for this simplification is increased conservativeness. This occurs because a *single* operator or relation $\bar{G}$ is replaced by a *set* of relations Cone(C, R). Many, perhaps most, of the input-output pairs in Cone(C, R) do not ever occur in the actual system $\bar{G}$, but since these input-output pairs must be implicitly included in any further analysis, the system's response may appear worse than it really is.

It is very important, therefore, to replace a system with a conic sector bound $\mathrm{Cone}(C, R)$ which is as tight as possible. Some systems cannot be adequately represented with conic sectors. Consider the nonlinear system $y = x^3$, where the input x ranges over the full set L_{2e}. This system has vanishingly small incremental gain $\frac{dy}{dx}$ for small inputs, but its gain grows without limit as the input x increases. No matter what center C is chosen, the radius R would have to have infinite gain to encompass the entire set of input-output pairs (x, y). Therefore this system is not suitable for conic sector analysis.

A sampled-data system has a time-varying response, so it is desirable to model it using conic sectors in order to simplify the analysis. This thesis outlines ways to find tight conic sector bounds for sampled-data systems, so that conic sector techniques may be used without introducing unnecessary conservativeness.

2.4 Conic Sector Stability Criterion

Consider the feedback system in Figure 2.4-1. Two relations $\bar{G}$ and $\bar{H}$ are connected in a positive feedback loop with external inputs u_1 and u_2 entering additively. The system equations are given by:

$$e_1 = u_1 + y_2 \qquad e_2 = u_2 + y_1$$

$$(e_1, y_1) \in \bar{G} \qquad (e_2, y_2) \in \bar{H}$$

Combine the two inputs and two outputs into column vectors $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. We define this feedback system to be *closed-loop stable* if there exists some finite $k > 0$ such that

$$\|y\|_\tau \leq k\|u\|_\tau \quad \forall\, \tau \geq 0$$

This definition of stability corresponds to "finite-gain stability" in [16]. It means that the outputs of $\bar{G}$ and $\bar{H}$ must have bounded energy for all inputs u_1 and u_2 of

bounded energy. The following test can determine when such a feedback system is stable:

Conic Sector Stability Criterion: Let C and R be LTI operators. If $\bar{G}$ is strictly inside Cone(C,R) and $\bar{H}^I$ is outside Cone(C,R), then the system in Figure 2.4-1 is closed-loop stable.

The proof is given in [16]. The Conic Sector Stability Criterion is a special case of the "sector stability theorem" of Safonov.

Note that this is only a sufficient condition for stability, not a necessary one; the system could be closed-loop stable without satisfying this cone separation condition for a given Cone(C,R). Nevertheless, this basic stability criterion has great utility and will be used extensively throughout this thesis.

The stability criterion as written above makes no assumptions for $\bar{G}$ and $\bar{H}$ other than that one must be inside a cone, and the inverse of the other outside the same cone. In the following pages, the criterion will always be used when $\bar{H}$ is known to be an LTI operator. We can take advantage of this to derive a version of the criterion that is more easily used — Safonov and Athans's Multiloop Circle Criterion [17]:

Multiloop Circle Criterion. Let $\bar{G}$ be a relation mapping L_{2e} to L_{2e} that is strictly inside Cone(C,R). Assume further that R^{-1} exists and is stable. Let $\bar{H}$ be an LTI operator with transfer function $H(j\omega)$. Then the system in Figure 2.4-1 is closed-loop stable if it is stable when $\bar{G}$ is replaced by its cone center C, and if

$$\|RH(I-CH)^{-1}\|_\infty \leq 1$$

Proof. Define the relation $\hat{G} = \bar{G} - C$ and redraw the system as in Figure 2.4-2. Then $\hat{G}$ is strictly inside Cone$(0,R)$. This is easily shown: since $\bar{G}$ is strictly inside Cone(C,R),

$$\|y_1 - Ce_1\|_\tau < \|Re_1\|_\tau \quad \forall\, \tau > 0, \quad \forall \text{ nonzero } e_1 \in L_{2e}$$

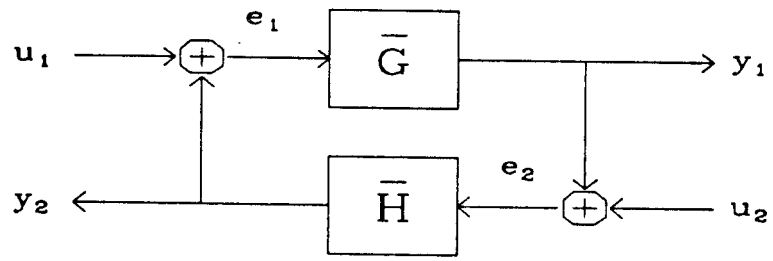

Fig. 2.4-1: A basic feedback loop. If $\bar{G}$ is strictly inside Cone(C, R) and $\bar{H}^I$ is outside Cone(C, R), then this system is stable.

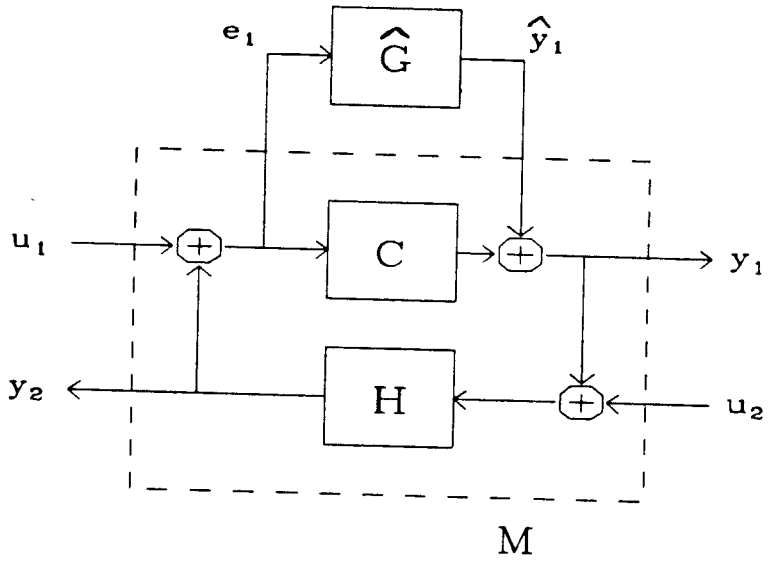

Fig. 2.4-2: $\bar{G}$ is replaced by $\hat{G} + C$. $\hat{G} = \bar{G} - C$ is inside Cone$(0, R)$. H is an LTI operator.

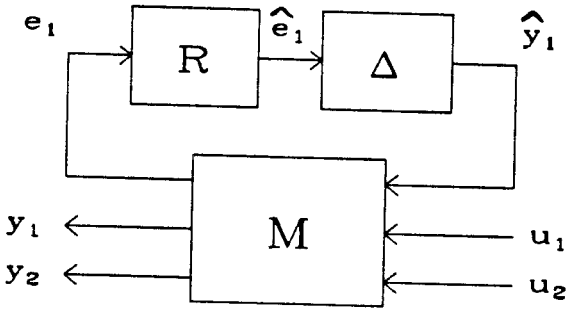

Fig. 2.4-3: $\hat{G}$ is replaced by $\Delta R = \hat{G}$. Δ is inside Cone$(0, I)$.

$$\hat{u}_1 = RM_{12}u_1 + RM_{13}u_2$$

$$\hat{u}_2 = 0$$

$$\hat{y}_2 = RM_{11}\hat{y}_1 = RH(I - CH)^{-1}\hat{y}_1$$

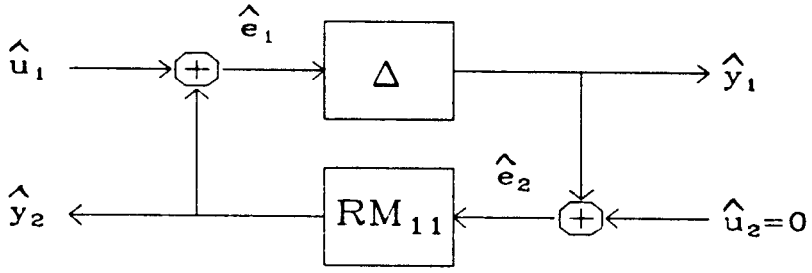

Fig. 2.4-4: The feedback loop containing Δ. This has the same form as Fig. 2.4-1.

by definition. The output of $\hat{G}$ is $\hat{y} = y_1 - Ce_1$; so $\|\hat{y}\|_\tau < \|Re_1\|_\tau$ by direct substitution. Therefore $\hat{G}$ is strictly inside Cone$(0, R)$.

Since C and $\bar{H}$ are LTI they may be lumped together into a single LTI operator M. Separate $\hat{G}$ from the three-input, three-output system in the dashed box and call the new system M. Its input-output equations are given by

$$\begin{bmatrix} e_1 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} H(I-CH)^{-1} & (I-HC)^{-1} & H(I-CH)^{-1} \\ (I-CH)^{-1} & C(I-HC)^{-1} & CH(I-CH)^{-1} \\ H(I-CH)^{-1} & HC(I-CH)^{-1} & H(I-CH)^{-1} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ u_1 \\ u_2 \end{bmatrix} = M \begin{bmatrix} \hat{y}_1 \\ u_1 \\ u_2 \end{bmatrix}$$

Since the feedback interconnection of C and H is stable by assumption, it can be shown using basic LTI feedback theory that all nine elements of M are stable.

Now let the relation Δ be defined by $\hat{G} = \Delta R$. We can show that Δ is strictly inside Cone$(0, I)$:

Let the input to Δ to be $\hat{e}_1$. The fact that $\hat{G}$ is strictly inside Cone$(0, R)$ implies that

$$\|\Delta\hat{e}_1\|_\tau = \|\hat{G}R^{-1}\hat{e}_1\|_\tau < \|RR^{-1}\hat{e}_1\|_\tau = \|I\hat{e}_1\|_\tau$$

for all $\tau > 0$, and for all nonzero $\hat{e}_1 \in L_{2e}$. Therefore Δ is strictly inside Cone$(0, I)$.

The system is redrawn in Figure 2.4-3. $\hat{G}$ is split into two parts R and Δ, and C and H are lumped into the stable system M. If $\hat{e}_1$ and $\hat{y}_1$ are the input and output, respectively, of Δ, $\hat{e}_1$ can be written as

$$\hat{e}_1 = RM_{11}\hat{y}_1 + RM_{12}u_1 + RM_{13}u_2$$

By defining $\hat{y}_2 \equiv RM_{11}\hat{y}_1$ and $\hat{u}_1 \equiv RM_{12}u_1 + RM_{13}u_2$, the expression reduces to $\hat{e}_1 = \hat{y}_2 + \hat{u}_1$ and the system can be redrawn as in Figure 2.4-4. This system has the same form as the basic feedback system in Figure 2.4-1, with Δ replacing $\bar{G}$ and $RM_{11} = RH(I-CH)^{-1}$ replacing $\bar{H}$. The external input $\hat{u}_2$ is identically zero.

Recall that Δ is strictly inside Cone$(0, I)$, so the system in Figure 2.4-4 is closed-loop stable if $(RH(I-CH)^{-1})^I$ is outside Cone$(0, I)$. Using the result at the end of

Section 2.2 involving the ∞-norm, and invoking the assumption that $H(I-CH)^{-1}$ is stable:

$$\begin{aligned}
& \|RH(I-CH)^{-1}\|_\infty \leq 1 \\
\Rightarrow \quad & \|\hat{y}_2\| \leq \|\hat{e}_2\| \quad \forall\, \hat{e}_2 \in L_2 \\
\Rightarrow \quad & \|\hat{y}\|_\tau \leq \|\hat{e}_2\|_\tau \quad \forall\, \tau > 0, \quad \forall\, \hat{e}_2 \in L_{2e} \\
\Rightarrow \quad & RH(I-CH)^{-1} \text{ is inside Cone}(0, I) \\
\Rightarrow \quad & (RH(I-CH)^{-1})^I \text{ is outside Cone}(0, I) \\
\Rightarrow \quad & \text{the system in Figure 2.4-4 is closed-loop stable.}
\end{aligned}$$

At this point, we know the system is stable with respect to the fictitious inputs and outputs $\hat{u}$ and $\hat{y}$. To complete the proof, we need to show that it is stable for the original inputs and outputs u, y. Recall that

$$\begin{aligned}
\hat{u}_1 &= RM_{12}u_1 + RM_{13}u_2 \\
\hat{u}_2 &= 0
\end{aligned}$$

and that RM_{12} and RM_{13} are stable. This implies that $\exists\, k_1 > 0$ such that

$$\|\hat{u}\|_\tau \leq k_1\|u\|_\tau \quad \forall\, \tau \geq 0$$

Now observe that the equations

$$\begin{aligned}
y_1 &= M_{21}\hat{y}_1 + M_{22}u_1 + M_{23}u_2 \\
y_2 &= M_{31}\hat{y}_1 + M_{32}u_1 + M_{33}u_3
\end{aligned}$$

imply that $\exists\, k_2, k_3 > 0$ such that

$$\|y\|_\tau \leq k_2\|\hat{y}\|_\tau + k_3\|u\|_\tau \quad \forall\, \tau \geq 0$$

since all the elements of M are stable. The system in Figure 2.4-4 is closed-loop stable, so by definition, $\exists\, k_4 > 0$ such that

$$\|\hat{y}\|_\tau \leq k_4\|\hat{u}\|_\tau \quad \forall\, \tau \geq 0$$

Therefore for all $\tau \geq 0$,

$$\begin{aligned} \|y\|_\tau &\leq k_2\|\hat{y}\|_\tau + k_3\|u\|_\tau \\ &\leq k_2 k_4\|\hat{u}\|_\tau + k_3\|u\|_\tau \\ &\leq (k_4 k_2 k_1 + k_3)\|u\|_\tau \end{aligned}$$

and the original system is closed-loop stable. This completes the proof. This proof is quite different from the proof in [17]. It was presented here in order to illustrate two analysis steps which will be used extensively in the new conic sector methods of Chapter 3. These are (a) subtracting the cone center C from an operator inside $\text{Cone}(C, R)$ to produce a new operator inside $\text{Cone}(0, R)$; and (b) factoring R from the input of an operator inside $\text{Cone}(0, R)$ to produce a new operator inside $\text{Cone}(0, I)$.

A useful interconnection rule for conic sectors and LTI operators was illustrated above when Δ was extracted from G. In general, for any relation G defined on $L_{2e} \times L_{2e}$ and any stable LTI operator F with stable inverse, we can cascade F with G at the input of G, resulting in a relation GF defined on $L_{2e} \times L_{2e}$. With these assumptions, the following rule holds:

$$G \text{ is inside } \text{Cone}(C, R) \quad \Leftrightarrow \quad GF \text{ is inside } \text{Cone}(CF, RF).$$

Proof. ($\Rightarrow$:) Define $x = Fe$. Then

$$\|y - Cx\|_\tau \leq \|Rx\|_\tau \quad \forall\, \tau > 0, \quad \forall\, (x, y) \in G \quad \text{(by assumption)}$$

$$\Rightarrow \quad \|y - CFe\|_\tau \leq \|RFe\|_\tau \quad \forall\, \tau > 0, \quad \forall\, (e, y) \in GF$$

$$\Rightarrow \quad GF \text{ is inside } \text{Cone}(CF, RF).$$

($\Leftarrow$:) Substitute F^{-1} for F.

The rule also holds if *strictly inside, outside*, or *strictly outside* is substituted for the word *inside*. This rule will be used several times in Chapter 3, usually in a way similar to its use above, where the cone bound $\text{Cone}(0, R)$ was replaced with the simpler cone bound $\text{Cone}(0, I)$.

2.5 The Structured Singular Value μ

Consider the feedback system in Figure 2.5-1, in which the two operators M and Δ are constant complex-valued matrices. The system is considered well-posed if the outputs e_1 and e_2 are defined for any finite inputs u_1 and u_2. The outputs are given by

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{pmatrix} (I-\Delta M)^{-1} & \Delta(I-M\Delta)^{-1} \\ M(I-\Delta M)^{-1} & (I-M\Delta)^{-1} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

This transfer matrix is defined as long as the two matrix inverses $(I-\Delta M)^{-1}$ and $(I-M\Delta)^{-1}$ exist; or equivalently, as long as $\det(I-\Delta M) \neq 0$ and $\det(I-M\Delta) \neq 0$. It is sufficient to check only one of these determinants, since if one of them is zero, so is the other. This is due to the well-established identity $\det(I-M\Delta) = \det(I-\Delta M)$.

Now suppose M is fixed, but Δ is a member of a set $\mathcal{X}$ of matrices, and we wish to find what bounds can be put on $\mathcal{X}$ so that this system is well-posed for all $\Delta \in \mathcal{X}$. This is analogous to the robust stability problem when M and Δ are transfer functions. If all that is known about Δ is its maximum gain, we can give a bound on $\bar{\sigma}(\Delta)$ which insures well-posedness:

$$\det(I-M\Delta) \neq 0 \quad \forall\, \Delta \in \{\Delta \mid \bar{\sigma}(\Delta) < \alpha\} \quad \Leftrightarrow \quad \bar{\sigma}(M) \leq 1/\alpha$$

where $\alpha > 0$ is a scalar. Since we know nothing about the internal structure of the matrix Δ, it is called an *unstructured uncertainty* block. Now consider the following case in which a great deal is known about the structure of Δ:

$$\det(I-M\Delta) \neq 0 \quad \forall\, \Delta \in \{\lambda I \mid \lambda \in C,\ |\lambda| < \alpha\} \quad \Leftrightarrow \quad \rho(M) \leq 1/\alpha$$

where $\rho(M) = \max_i |\lambda_i(M)|$ and $\alpha > 0$ is a scalar. In this case, the only unknown fact about Δ is the value of the scalar λ which appears in each diagonal element.

The two results above allow us to analyze the well-posedness of the system in Figure 2.5-1 in the face of uncertain Δ, but only in two rather extreme cases:

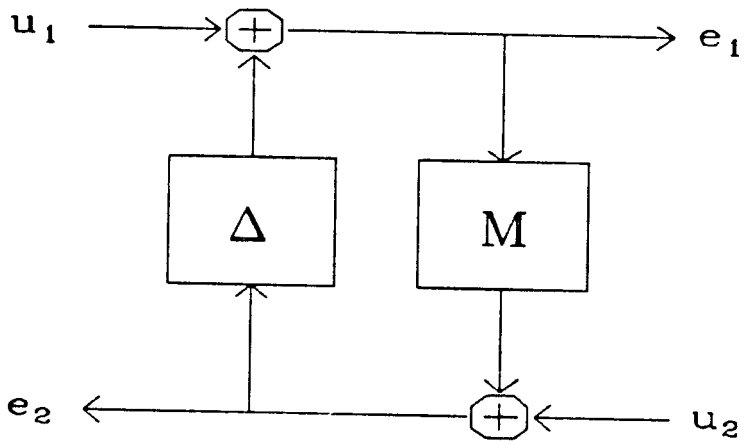

Fig. 2.5-1: A feedback loop with constant M, Δ. The system is well posed if $\det(I - M\Delta) \neq 0$.

(1) unstructured uncertainty, in which the maximum gain of Δ is all that is known about it; and (2) the case in which Δ is known to be a scalar times the identity matrix. However, one's knowledge of Δ might well lie between these two extremes; for example, a 2 by 2 matrix $\Delta = [\delta_{ij}]$ might be known to have $\delta_{21} = 0$ and $|\delta_{ij}| < 1$ for the other values of i and j. A method is needed to handle such cases. It is sufficient for the method to handle only *block diagonal* Δ, because a problem such as the one above can always be rearranged into block diagonal form:

$$\text{let} \quad \Delta = \begin{bmatrix} \delta_1 & \delta_2 \\ 0 & \delta_3 \end{bmatrix} \qquad |\delta_i| < 1$$

$$\text{define} \quad \tilde{\Delta} = \begin{bmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_2 & 0 \\ 0 & 0 & \delta_3 \end{bmatrix}$$

and note that

$$\Delta = \begin{bmatrix} \delta_1 & \delta_2 \\ 0 & \delta_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_2 & 0 \\ 0 & 0 & \delta_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = U\tilde{\Delta}V$$

Therefore the matrices U and V transform the original problem into one with a block-diagonal $\tilde{\Delta}$:

$$\det(I - M\Delta) = 0 \quad \Leftrightarrow \quad \det(I - VMU\tilde{\Delta}) = 0$$

Proof:

$$\det(I - M\Delta) = \det(I - MU\tilde{\Delta}V) = 0$$

$$\Leftrightarrow \quad \exists\, x \neq 0 \text{ such that } MU\tilde{\Delta}Vx = x$$

$$\Leftrightarrow \quad \exists\, (Vx) \neq 0 \text{ such that } VMU\tilde{\Delta}(Vx) = (Vx)$$

$$\Leftrightarrow \quad \exists\, y \neq 0 \text{ such that } VMU\tilde{\Delta}y = y$$

$$\Leftrightarrow \quad \det(I - VMU\tilde{\Delta}) = 0$$

This block-diagonal perturbation problem can be handled with the Structured Singular Value μ [5],[6]. Define a function $\mu(M)$ such that

$$\det(I - M\Delta) \neq 0 \quad \forall\, \Delta \in \mathcal{X}_\alpha \quad \Leftrightarrow \quad \mu(M) < 1/\alpha$$

where the scalar $\alpha > 0$ and

$$\mathcal{X}_\alpha \equiv \{\text{diag}(\Delta_1, \Delta_2, \ldots, \Delta_n) \mid \bar{\sigma}(\Delta_i) \leq \alpha \text{ for } i = 1 \text{ to } n\}$$

is the set of block-diagonal matrices with gain less than or equal to α. Each square block Δ_i has an asuumed dimension k_i. Observe that $\mu(M)$ is a function not only of the matrix M, but also of the assumed structure of Δ given by the set $\mathcal{X}_\alpha$; that is, the number and size of the blocks Δ_i of which Δ is composed. Let

$$\mathcal{X}_\infty \equiv \{\text{diag}(\Delta_1, \Delta_2, \ldots, \Delta_n)\}$$

be the set of block-diagonal matrices Δ_i with $\dim(\Delta_i) = k_i$ and with no restriction on gain. Then an alternative definition of $\mu(M)$ is given in [5]:

$$\mu(M) = \begin{cases} 0 & \text{if no } \Delta \in \mathcal{X}_\infty \text{ solves} \\ & \det(I - M\Delta) = 0; \\ \dfrac{1}{\inf_{\Delta \in \mathcal{X}} \{\bar{\sigma}(\Delta) \mid \det(I - M\Delta) = 0\}} & \text{otherwise.} \end{cases}$$

In other words, $\mu(M)$ is the reciprocal of the gain of the smallest block-diagonal matrix $\Delta \in \mathcal{X}_\infty$ which causes the feedback system to be ill-posed. Compare this to the results given above for $\bar{\sigma}(M)$ and $\rho(M)$: (1) $\bar{\sigma}(M)$ is the reciprocal of the gain of the smallest unstructured matrix Δ which makes the system ill-posed; and (2) $\rho(M)$ is the reciprocal of the gain of the smallest "scalar times identity" matrix which makes the system ill-posed. Given that the case of block-diagonal Δ is intermediate between cases (1) and (2), it is not too surprising that for any structure $\mathcal{X}_\infty$,

$$\rho(M) \leq \mu(M) \leq \bar{\sigma}(M)$$

$\mu(M)$ can be thought of as a measure of the closeness of the feedback system to ill-posedness (and to instability when M and Δ are transfer functions, as we shall see) when the structure of Δ is known to be block-diagonal; hence the name *Structured* Singular Value. Furthermore, it is applicable to general structures for Δ

since, as was demonstrated above, non-diagonal structures can be rearranged into diagonal ones. μ-analysis is applicable to any LTI feedback system in which several parameters vary independently over a bounded range of values.

Here are some important properties of μ:

$$\mu(\alpha M) = |\alpha|\mu(M)$$

$$\mu(M) \leq \inf_{D \in \mathcal{D}} \bar{\sigma}(DMD^{-1})$$

$$\sup_{U \in \mathcal{U}} \rho(MU) \leq \mu(M)$$

where

$$\mathcal{D} \equiv \{D \mid D = \text{diag}(d_1 I, d_2 I, \ldots, d_n I),\ d_i \in R\}$$

$$\mathcal{U} \equiv \{U \mid U = \text{diag}(U_1, U_2, \ldots, U_n),\ U_i \text{ unitary}\}$$

and the dimensions of the diagonal subblocks $d_i I$ and U_i match those of the subblocks Δ_i in $\mathcal{X}_\infty$. The last two properties are especially important since they provide upper and lower bounds for μ. There is, as yet, no known general algorithm to compute μ exactly, so all μ software to date actually computes one of these bounds instead. Fortunately, the difference between the upper and lower bounds is usually quite small (less than 10 percent), so this does not cause much loss of accuracy. The upper bound is more frequently computed, since the optimization problem associated with it is convex; this guarantees that the global minimum will be found. Software to compute the upper bound generally runs faster than software for the lower bound. However, progress has been made recently [8] in finding faster algorithms to compute the lower bound.

2.6 Robustness Analysis for LTI Systems

2.6.1 Linear Fractional Transformations

Consider the feedback system in Figure 2.6-1, in which the (possibly MIMO) transfer function Δ is connected in a feedback arrangement with the MIMO transfer function $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$, a block 2 by 2 matrix. We will denote the resulting transfer function from u to y by the convenient notation

$$F_u(P, \Delta) \equiv P_{22} + P_{21}\Delta(I - P_{11}\Delta)^{-1}P_{12}$$

where the subscript u stands for "upper." A lower linear fractional transformation (LFT) can be defined similarly. In Figure 2.6-2, the 2-by-2 block system P is connected in a feedback arrangement with K, situated below it on the block diagram. Denote the resulting transfer function from u to y by

$$F_l(P, K) \equiv P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}$$

where the letter l stands for "lower."

2.6.2 Nominal Stability

Consider again the block diagram in Figure 2.6-1. We will generally think of P as the nominal, i.e., unperturbed plant, while Δ will represent the allowable perturbations or uncertainties in the plant. The word "plant" is used in a general fashion here, since it might well include the feedback compensator. The objective of this analysis is to determine the effects of model uncertainty on the closed-loop feedback system *after* a compensator has been designed and connected to the open-loop plant, so P will generally describe the nominal closed-loop system. Δ will usually be assumed to be stable ($\Delta \in H_\infty$), norm-bounded by 1, and to have block-

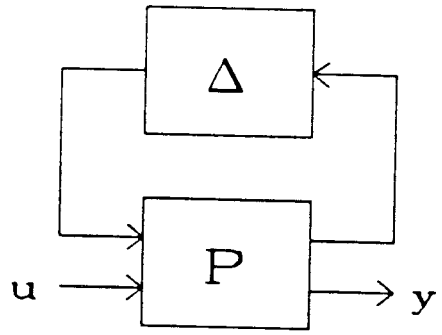

Fig. 2.6-1: An upper linear fractional transformation.
$y = F_u(P, \Delta)u = (P_{22} + P_{21}\Delta(I - P_{11}\Delta)^{-1}P_{12})u.$

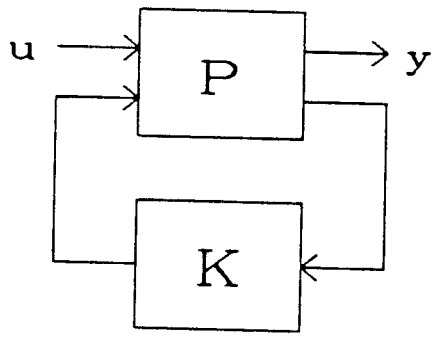

Fig. 2.6-2: A lower linear fractional transformation.
$y = F_l(P, K)u = (P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21})u.$

diagonal structure: $\Delta \in \mathcal{X}_1$ where

$$\mathcal{X}_1 \equiv \{\Delta \mid \Delta = \text{diag}(\Delta_1, \Delta_2, \ldots, \Delta_n) \quad \text{and} \quad \|\Delta_i\|_\infty \leq 1\}$$

As before, each Δ_i is assumed to be square, with $\dim(\Delta_i) = k_i$. This differs from the definition of $\mathcal{X}_\alpha$ in Section 2.5 only in that each Δ_i is now a frequency dependent transfer function, not just a constant matrix. These conditions may appear restrictive, but in fact they are very general: if the actual plant uncertainty $\tilde{\Delta}_i$ is not equal in magnitude to 1 at all frequencies, simply construct a frequency-dependent weighting function W_i such that $\bar{\sigma}(W_i\tilde{\Delta}_i) \leq 1$ for all frequencies, and absorb W_i into the nominal plant. If the actual plant uncertainty does not have a block-diagonal structure, we have already seen that the system can be rearranged to give it such a structure.

The input-output behavior of the perturbed system with a particular perturbation Δ is simply $F_u(P, \Delta)$. To determine if the nominal system is stable, check the stability of $F_u(P, \Delta)$ with $\Delta = 0$. But $F_u(P, 0) = P_{22}$; therefore

$$\text{The system is nominally stable} \quad \Leftrightarrow \quad P_{22}\text{is stable.}$$

2.6.3 Robust stability

A much more interesting question is whether this feedback system remains stable for all perturbations $\Delta \in \mathcal{X}_1$. To answer this question, we need to state exactly what "stable" means in this context. Consider Figure 2.6-3, in which disturbance inputs u_1 and u_2 have been added at both sides of Δ and the resulting "error" signals e_1 and e_2 have been brought outside the system. Define the feedback system $F_u(P, \Delta)$ to be *internally stable* if the transfer function from $\begin{pmatrix} u_1 \\ u_2 \\ u \end{pmatrix}$ to $\begin{pmatrix} e_1 \\ e_2 \\ y \end{pmatrix}$ is stable; i.e., analytic in the right-half complex plane C_+. The following important theorem from [6] establishes robust stability for this system:

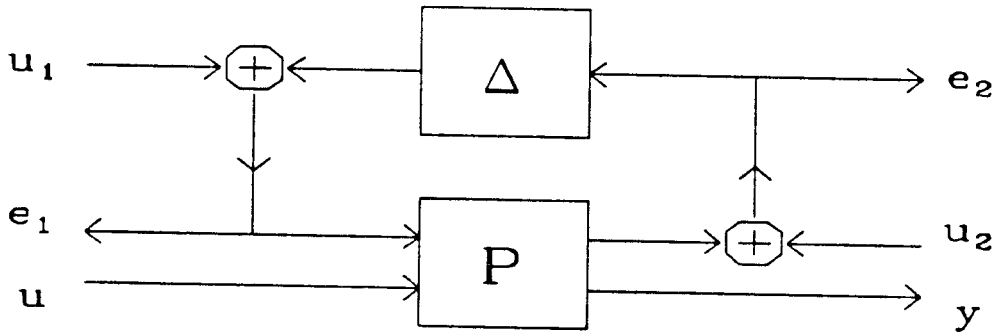

Fig. 2.6-3: A feedback loop with disturbance inputs u_1, u_2.

Robust Stability Theorem. If P is stable, $F_u(P, \Delta)$ is internally stable for all real-rational, stable $\Delta \in \mathcal{X}_1$ if and only if

$$\|\mu(P_{11})\|_\infty \leq 1.$$

The fact that this gives a necessary and sufficient condition for robust stability means that this test is completely nonconservative, as long as the chosen structure for Δ accurately reflects the possible variation in the closed-loop plant.

2.6.4 Nominal Performance

Good performance for a feedback regulator usually means that the change in the regulated output is small for large changes in the system input. In other words, the performance is good if the closed-loop transfer function is small at all frequencies. For other types of feedback systems, this may not be true; for instance, a closed-loop servo system is performing well if the output tracks the input over a wide range of frequencies. Good performance in this case means that the closed-loop transfer function closely matches a prescribed low-pass characteristic. In these cases, we can define an "error" output to be the algebraic difference between the actual output and the desired, or nominal, output. With the error output so defined, good performance is equivalent to a small gain at all frequencies for the transfer function to this fictitious output, as in the case of a regulator.

For robustness analysis, then, one can always define outputs which are the difference between the outputs of the actual system and the outputs of a "reference system" having the desired response. Frequency weights can easily be attached to these outputs and absorbed into the system model. After these two steps of subtracting the desired response and applying frequency weights, the performance criterion can always be put into the form

$$\text{satisfactory performance} \quad \Leftrightarrow \quad \|M\|_\infty \leq 1$$

where M is an appropriately defined transfer function. Define *nominal performance* to mean satisfactory performance for the unperturbed feedback system P in Figure 2.6-1, i.e., with $\Delta = 0$. Assume that the output y is a weighted error output as described above; then

$$\text{nominal performance} \quad \Leftrightarrow \quad \|P_{22}\|_\infty \leq 1$$

2.6.5 Robust Performance

Since P_{22} is an exactly known transfer function, determining whether the system meets nominal performance requirements, i.e., $\|P_{22}\|_\infty \leq 1$, is trivial; simply compute and plot the maximum singular value of $P_{22}(j\omega)$. The important question is whether it meets these requirements, and remains internally stable, for all allowable perturbations:

$$\begin{cases} (i) & \|F_u(P,\Delta)\|_\infty \leq 1 \quad \forall \text{ stable } \Delta \in \mathcal{X}_1, \text{ and} \\ (ii) & \|\mu(P_{11})\|_\infty \leq 1 \end{cases}$$

By the small gain theorem, this is equivalent to robust stability for the structure shown in Figure 2.6-4, in which another uncertainty block Δ_{n+1} has been connected from the lower output of P to its lower input:

$$\begin{cases} (i) & \|F_u(P,\Delta)\|_\infty \leq 1 \quad \forall \text{ stable } \Delta \in \mathcal{X}_1, \text{ and} \\ (ii) & \|\mu(P_{11})\|_\infty \leq 1 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} (i) & \det(I - F_u(P(j\omega), \Delta(j\omega))\Delta_{n+1}(j\omega)) \neq 0 \\ & \qquad \forall\, \omega \in R,\ \forall\, \Delta, \Delta_{n+1} \in H_\infty \text{ where } \Delta \in \mathcal{X}_1,\ \|\Delta_{n+1}\|_\infty \leq 1, \text{ and} \\ (ii) & \|\mu(P_{11})\|_\infty \leq 1 \end{cases}$$

$\Leftrightarrow$ the system in Figure 2.6-4 is robustly stable.

However, we can simply augment Δ with Δ_{n+1} as shown in Figure 2.6-5 to produce a new perturbation matrix $\tilde{\Delta} = \begin{bmatrix} \Delta & 0 \\ 0 & \Delta_{n+1} \end{bmatrix}$ which is also block-diagonal. Robust

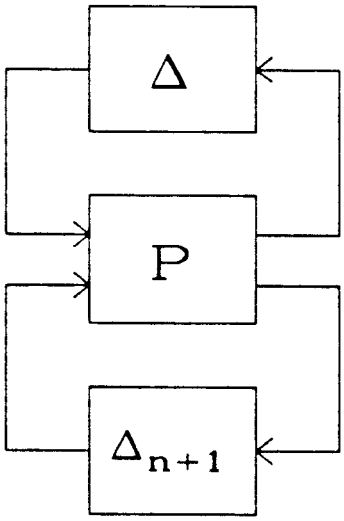

Fig. 2.6-4: A system with fictitious "performance block" Δ_{n+1}. Stable for $\|\Delta_{n+1}\|_\infty < 1$ iff $\|F_u(P,\Delta)\|_\infty < 1$.

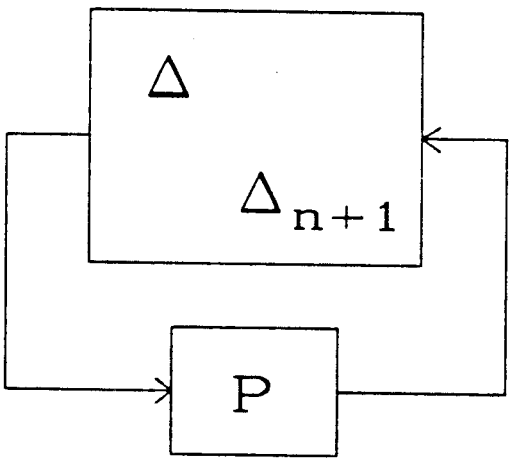

Fig. 2.6-5: Same as Fig. 2.6-4, but with Δ_{n+1} augmented with Δ. $\|F_u(P,\Delta)\|_\infty < 1$ iff $\|\mu(P)\|_\infty < 1$.

stability for Figure 2.6-5 is obviously equivalent to robust stability for Figure 2.6-4, since the blocks have only been rearranged. Robust stability for Figure 2.6-5 can be determined, as before, by computing μ at all frequencies, but now μ should be computed for the entire block 2-by-2 transfer function P. This is summed up in the following theorem from [6]:

Robust Performance Theorem. The system in Figure 2.6-1 meets robust performance requirements (i.e., it is robustly stable and satisfies $\|F_u(P,\Delta)\|_\infty \leq 1 \quad \forall$ stable $\Delta \in \mathcal{X}_1$, if and only if

$$\|\mu(P)\|_\infty \leq 1.$$

This powerful theorem shows that the Structured Singular Value μ is a very useful figure of merit for system robustness. By defining Δ-blocks and error outputs with frequency weights that accurately reflect model uncertainty and closed-loop performance requirements, a designer can determine if a system will meet its specifications under all allowed perturbations simply by computing and plotting μ. If $\mu < 1$ at all frequencies, the system is robust.

There is a controller synthesis method, based on recent H_∞ synthesis results, which automatically produces stabilizing controllers to minimize $\|\mu\|_\infty$. A discussion of it is beyond the scope of this thesis; it is described in [6] and [18].

2.7 Sampled-data Systems

2.7.1 Digital Filters: A Continuous-time Interpretation

A digital filter with z-transform $D(z)$ is usually thought of as an operator mapping an input *time sequence* $\{u_0, u_1, u_2, \ldots\}$ into an output time sequence $\{y_0, y_1, y_2, \ldots\}$. In this conceptual framework, the system inputs and outputs are considered to exist only at times t which are multiples of the sampling period T: $t = \{\ldots, -2T, -T, 0, T, 2T, \ldots\}$. An alternative viewpoint is to view $D(z)$ as a *continuous-time operator* with an impulse response that is an impulse train: $d(t) = \sum_k d_k \delta(t - kT)$ where $D(z) = \sum_k d_k z^{-k}$. With this definition of impulse response, the frequency response is readily recovered by taking the Laplace transform of $d(t)$ in the usual manner:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-st} d(t) dt &= \int_{-\infty}^{\infty} e^{-st} \sum_k d_k \delta(t - kT) dt \\ &= \sum_k d_k e^{-sTk} = D(e^{sT}) \end{aligned}$$

where the shorthand notation $\sum_k$ is used to represent $\sum_{k=-\infty}^{\infty}$.

The Laplace transform of $d(t)$ evaluated on the $j\omega$-axis is equal to the z-transform $D(z)$ evaluated on the unit circle, so these two ways of looking at the action of the digital filter $D(z)$ are equally valid.

Usually, the input to a digital filter is restricted to be an impulse train, but the development above shows that we can define its output when the input is a continuous-time waveform. We treat $D(z) = D(e^{j\omega T})$ as a continuous-time LTI operator, so the Fourier transform of its output is simply the product of $D(e^{j\omega T})$ with the Fourier transform of the input. $D(e^{j\omega T})$ can be thought of as a linear interconnection of weighted time-delay elements $d_k e^{-skT}$; the delay element corresponding to d_k has a time delay of kT seconds.

2.7.2 z-Transform Loop Stability Analysis

Figure 2.7-1 shows a sampled-data feedback system. The sampler (treated as an impulse modulator) converts the analog output $y(t)$ of the plant $G(s)$ into an impulse train $\sum_k y(kT)\delta(t-kT)$. The impulse train is processed by the digital filter $D(z) = D(e^{j\omega T})$, and the result is converted back into an analog signal by the hold operator $H_0^T(s)$. The notation H_0^T indicates that the hold is a zero-order hold (ZOH) with length T, defined by $H_0^T(s) = (1-e^{-sT})/s$. The impulse response of $H_0^T(s)$ is a rectangular pulse with height 1 and a duration of T seconds. The effect of $H_0^T(s)$ is to convert an impulse train with pulse spacing T into a stair-step analog waveform. This waveform is the input to the analog plant $G(s)$.

The output of a standard digital-to-analog converter (DAC) is a stair-step waveform of this type. Therefore most practical sampled-data systems implicitly include zero-order holds. More complicated holds, such as the first-order hold (FOH) $H_1^T(s) \equiv (1-e^{-sT})^2(s+1/T)/s^2$ and the slewer hold $H_S^T(s) \equiv (1-e^{-sT})^2/Ts^2$, are occasionally used in practice to achieve a better match between the DAC output and some desired smooth analog signal, or to make the DAC output continuous for systems which are sensitive to step inputs.

If the analog plant transfer function $G(s)$ is *exactly* known, then the combination of the hold, the plant, and the sampler can be viewed as a discrete-time LTI operator, and standard z-transform techniques may be used to determine stability. When an impulse $\delta(t)$ is applied to the ZOH's input, the resulting impulse train at the output of the sampler is the impulse response of this discrete-time equivalent system. Let $g(t)$ be the impulse response of $G(s)$. Then $y(t) = g(t) * h(t)$ is the input to the sampler, where $h(t)$ is the ZOH impulse response. If $G(s)$ has a state-space representation, then $y(kT)$ may be derived as follows:

$$G(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

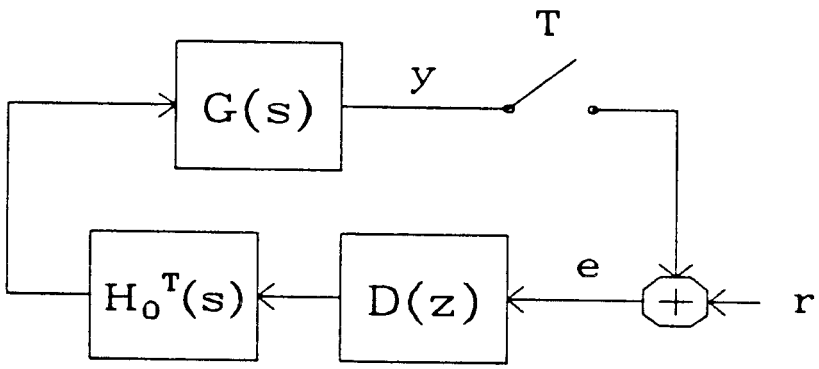

Fig. 2.7-1: A sampled-data feedback system.

So

$$g(t) = Ce^{At}B + D\delta(t)$$

$h(t)$ is given by

$$h(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

So $y(t) = g(t) * h(t)$; since $h(t) = 0$ for $t \ge 1$, consider $y(t)$ for $t \ge 1$:

$$y(t) = \int_0^t g(t-\tau)h(\tau)d\tau = \int_0^T g(t-\tau)d\tau$$

$$\int_0^T Ce^{A(t-\tau)}Bd\tau = Ce^{A(t-T)}\int_0^T e^{A(T-\tau)}Bd\tau = Ce^{A(t-T)}\int_0^T e^{A\tau}Bd\tau$$

Now define Γ and Φ by

$$\Gamma = \int_0^T e^{A\tau}Bd\tau \qquad \Phi = e^{AT}$$

For positive sample times $t = kT$ we get the sampled impulse response

$$y(kT) = \begin{cases} D, & k = 0 \\ C\Phi^{k-1}\Gamma, & k \ge 1 \end{cases}$$

This impulse response has the state-space representation

$$Y(z) = \left[\begin{array}{c|c} \Phi & \Gamma \\ \hline C & D \end{array}\right]$$

Both Φ and Γ may be computed together in one step, using a matrix exponential routine and the following identity:

$$\exp\left(\begin{bmatrix} AT & BT \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} e^{AT} & \int_0^T eA\tau B d\tau \\ 0 & 1 \end{bmatrix}$$

Given this z-transform for the discretized plant, it is simple to determine stability of the loop. The discrete-time closed-loop transfer function from r to e is $(I - Y(z)D(z))^{-1}$; simply compute the state-space representation for this function using the rules of Section 2.2.1, then compute the system's poles. If all poles are inside

the unit circle, the system is stable; if not, the system is unstable. Alternatively, we could perform the discrete Nyquist test on the loop gain $Y(z)D(z)$.

The problem with this approach is that it assumes the plant transfer function $G(s)$ is *exactly* known. This is never the case in practice. For any physical system, the best we can do is find a $G(s)$ which closely approximates the system response, and to find some bounds at each frequency on the error between this model and the actual system. If the system is intended to operate over a wide range of temperature, air pressure, component tolerances, etc., any meaningful plant model should include these uncertainty bounds as an important part of the model.

This issue of system robustness with respect to plant uncertainty is the motivation for using conic sector analysis on sampled-data systems. Since the uncertainty is characterized in the analog domain, it is desirable to treat all signals and all operators in the system as analog entities. The price paid for this treatment is that in the analog domain, the system is not LTI. The techniques described in Chapter 3 let us work around this difficulty without sacrificing accuracy.

2.7.3 Notation for Sampled Signals and Systems

There is a useful notation, due to Richard Whitbeck [26], that describes both sampled signals and the transfer functions of discretized analog plants. If $u(t)$ is a continuous-time signal (possibly vector-valued), then define

$$(u)^T = \sum_{k=0}^{\infty} u(kT)\delta(t - kT).$$

So $(u)^T$ is the output of a sampler (impulse modulator) of period T when its input is $u(t)$. The superscript indicates the period of the sampler; this makes the notation well suited for multi-rate systems, where all samplers do not run at the same speed.

In a similar fashion, define

$$G^T(z) = \sum_{k=0}^{\infty} g(kT)z^{-k}$$

where $g(t)$ is the (possibly matrix-valued) impulse response of an analog plant with transfer function $G(s)$. $G^T(z)$ is the z-transform of the discretized plant $G(s)$. Again, the superscript indicates the sampling period.

Sometimes it is more convenient to write $(u)^T$ without parentheses, as u^T. To avoid confusion, lower-case letters will generally indicate signals, and upper-case letters will indicate transfer functions.

2.7.4 Hybrid Operators

Consider the sampled-data operator in Figure 2.7-2. An analog input signal $u(t)$ passes through an LTI prefilter $F(s)$ and the result $(Fu)(t)$ is sampled at a period of T seconds per sample. The sampler is modelled here, and throughout this thesis, as an impulse modulator: its output is a weighted string of δ-functions. This impulse train is denoted $(Fu)^T(t) = \sum_k \delta(t - kT)(Fu)(t)$, following Whitbeck's notation. $(Fu)^T$ passes through the discrete-time operator $D(z) = D(e^{j\omega T})$, which is a conventional digital filter. The output $(DFu)^T(t)$ of $D(e^{j\omega T})$ feeds into the LTI hold filter $H(s)$ and the resulting system output is $y(t) = h(t) * (D(Fu)^T)$, the linear convolution of the impulse train $(DFu)^T(t)$ with the impulse response $h(t)$ of $H(s)$. This type of system is called a *hybrid* operator, since it is a combination of continuous-time and discrete-time systems.

Since all the components of this hybrid operator are linear, so is the hybrid itself. However, due to the action of the sampler, the hybrid is not time-invariant; therefore no transfer function is defined for this system, and conventional frequency-domain analysis techniques are useless. There is, however, a useful property of hybrid operators which can be exploited: they are *periodically* linear time-varying.

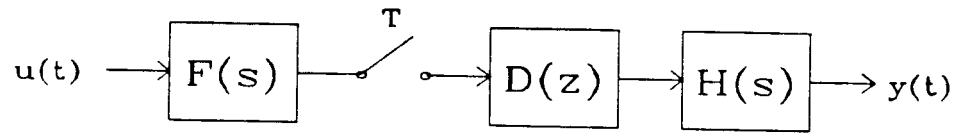

Fig. 2.7-2: A hybrid operator.

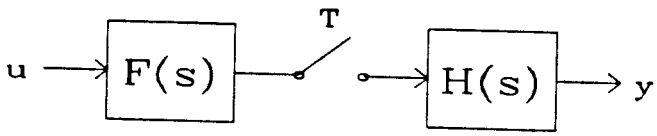

Fig. 2.7-3: A hybrid operator with $D(e^{sT})$ lumped together with the hold $H(s)$.

This means that, if $g(t,\tau)$ is the time response of the system to an input impulse at time τ:

$$g(t+T,\tau+T) \;=\; g(t,\tau) \quad \forall\, t,\tau$$

where T is the period of the system (in this case, the period of the sampler). This can easily be shown:

For simplicity, lump the digital filter $D(e^{j\omega T})$ together with the hold $H(s)$; since both operators are LTI, this is legal. This results in the hybrid of Figure 2.7-3. Compute the output $y(t) = h(t) * (Fu)^T$ when the input is a delta function at time τ: $u(t) = \delta(t-\tau)$. Let $f(t)$ and $h(t)$ denote the impulse responses of $F(s)$ and $H(s)$, respectively.

$$\begin{aligned}
(Fu)(t) &= \int_{-\infty}^{\infty} f(t-s)\delta(s-\tau)ds \;=\; f(t-\tau) \\
(Fu)^*(t) &= \sum_k \delta(t-kT)(Fu)(t) \;=\; \sum_k \delta(t-kT)f(kT-\tau) \\
h(t) * (Fu)^T(t) &= \int_{-\infty}^{\infty} h(t-s)\sum_k \delta(s-kT)f(kT-\tau)ds \\
&= \sum_k h(t-kT)f(kT-\tau) \;=\; g(t,\tau)
\end{aligned}$$

Now to show that the system is PLTV with period T, evaluate $g(t+T,\tau+T)$:

$$\begin{aligned}
g(t+T,\tau+T) &= \sum_k h(t+T-kT)f(kT-\tau-T) \\
&= \sum_k h(t-(k-1)T)f((k-1)T-\tau) \\
&= \sum_k h(t-kT)f(kT-\tau) \\
&= g(t,\tau)
\end{aligned}$$

2.7.5 A Property of PLTV Operators

Any two SISO LTI operators D and G are commutative; i.e., $DGu = GDu$ for all inputs u. If D or G is PLTV, however, this is no longer true in general. A condition which makes them commutative is that the LTI operator $D(j\omega)$ be periodic in ω

with period $2\pi/T$ where T is the period in seconds of the PLTV operator G. This property is easily proven, and will be exploited later in this thesis.

Let $g(t,\tau)$ be the response of the PLTV operator G to an impulse at time τ. Assume that G is periodic with period T: $g(t+T,\tau+T) = g(t,\tau)$. Assume that $D(j\omega)$ is the frequency response of the LTI operator D and that $D(j(\omega+2\pi/T)) = D(j\omega)$ for all real ω.

Then the impulse response $d(t)$ of D can be found by taking the Fourier series of $D(j\omega)$:

$$\begin{aligned} d(t) &= \sum_k d_k \delta(t-kT) \quad \text{where} \\ d_k &= \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} e^{j\omega kT} D(j\omega)\, d\omega \end{aligned}$$

Note that $d(t)$ is an impulse train, so $D(j\omega)$ can be thought of as the frequency response of a digital filter. Now for any input $u(t)$:

$$\begin{aligned} (DGu)(t) &= d(t) * ((Gu)(t)) = \int_{-\infty}^{\infty} d(t-s)((Gu)(s))ds \\ &= \int_{-\infty}^{\infty} d(t-s)\int_{-\infty}^{\infty} g(s,\tau)u(\tau)d\tau ds \\ &= \int_{-\infty}^{\infty} \sum_k d_k \delta(t-s-kT)\int_{-\infty}^{\infty} g(s,\tau)u(\tau)d\tau ds \\ &= \sum_k d_k \int_{-\infty}^{\infty} g(t-kT,\tau)u(\tau)d\tau \end{aligned}$$

and

$$\begin{aligned} (GDu)(t) &= \int_{-\infty}^{\infty} g(t,\tau)((Du)(\tau))d\tau \\ &= \int_{-\infty}^{\infty} g(t,\tau)\int_{-\infty}^{\infty} d(\tau-s)u(s)ds d\tau \\ &= \int_{-\infty}^{\infty} g(t,\tau)\int_{-\infty}^{\infty} \sum_k d_k \delta(\tau-s-kT)u(s)ds d\tau \\ &= \sum_k d_k \int_{-\infty}^{\infty} g(t,\tau)u(\tau-kT)d\tau \\ &= \sum_k d_k \int_{-\infty}^{\infty} g(t,\tau+kT)u(\tau)d\tau \\ &= \sum_k d_k \int_{-\infty}^{\infty} g(t-kT,\tau)u(\tau)d\tau \\ &= (DGu)(t) \end{aligned}$$

As long as $D(j\omega) \neq 0$ for any real ω, this identity can be expressed in another way:

$$Gu = DGD^{-1}u$$

for all inputs u. This is easily shown: $DGD^{-1}u = GDD^{-1}u = Gu$. This is the form in which this identity will be used most often.

2.8 Multirate Analysis: Kranc Operators in State Space

One of the main results of this thesis is a nonconservative robustness analysis technique for multi-rate hybrid systems. In order to have a standard for comparison, we need a technique to compute the robustness margins *exactly* in some special cases; then we can compare the robustness margin computed with the new conic sector method to the exact, known margin in order to estimate the conservativeness of the new technique. A good method for doing this is Kranc vector switch decomposition in state space, previously reported in [22],[21].

2.8.1 State-space Realizations for Kranc Operators

Consider the multirate system in Figure 2.8-1. An analog plant $G(s)$ (SISO or MIMO) is fed by the output of a sampler running at period T/m, and the output of $G(s)$ is sampled at period T/n, where n and m are integers. The two sampling periods are thereby related by a rational number; this assumption is required for the Kranc operator approach. The samplers are modelled by impulse modulators, as usual. Since the sampled input sequence $(u)^{T/m}$ and the sampled output sequence $(y)^{T/n}$ run at different rates, the relationship between them cannot be shift-invariant; therefore there can exist no transfer function relating these two sequences directly.

Figure 2.8-2 shows the same system in an equivalent form: each sampler has been replaced by a multivariable sampler with period T, with the LTI operators E^{m+} and E^{m-}, or E^{n+} and E^{n-}, at its input and output. E^{n+} and E^{n-} are defined as follows:

$$E^{n+} = \begin{bmatrix} I \\ e^{sT/n}I \\ e^{2sT/n}I \\ \vdots \\ e^{(n-1)sT/n}I \end{bmatrix}$$

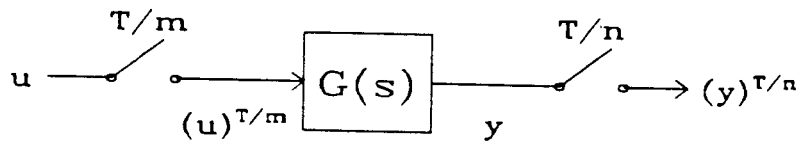

Fig. 2.8-1: A multirate system. m and n are integers.

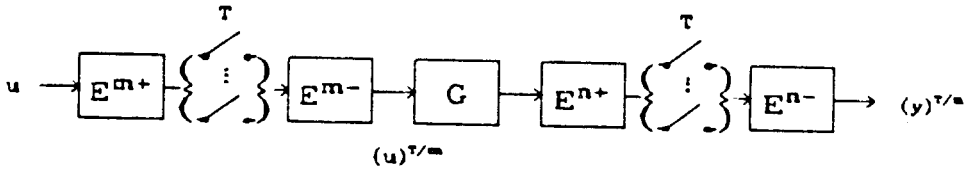

Fig. 2.8-2: Same system with samplers converted to period T by decomposition and recombination operators.

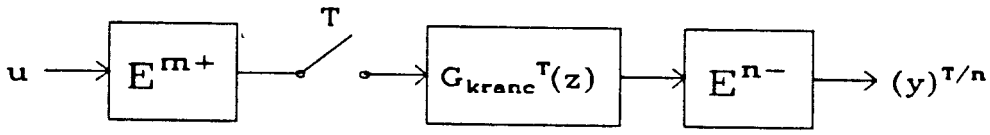

Fig. 2.8-3: Same system with the discrete-time LTI operator between the samplers replaced by its Kranc operator $G^T_{Kranc}(z)$. The sampler is MIMO.

$$E^{n-} = \begin{bmatrix} I & e^{-sT/n}I & e^{-2sT/n}I & \dots & e^{-(n-1)sT/n}I \end{bmatrix}$$

where each I is an identity matrix of appropriate dimension. E^{n+} has the effect of creating multiple time-advanced copies of its input signal. When the "vectorized" signal $E^{n+}u$ is sampled at period T, the resulting vector-valued sequence $(E^{n+}u)^T$ contains the same values as the original signal sampled at the faster period T/n: $(u)^{T/n}$. The process can be thought of as taking a "frame" of n consecutive samples from the fast-rate sequence $(u)^{T/n}$ and arranging them into a single vector-valued sample of the slow-rate sequence $(E^{n+}u)^T$.

The operator E^{n-} recombines such a slow-rate vector sequence back into a fast-rate sequence. It applies increasing amounts of time delay to its n inputs to cancel the increasing amounts of time advance that were applied to these signals by E^{n+}. E^{n+} and E^{n-} are called decomposition and recomposition operators, respectively.

Note that since both samplers in Figure 2.8-2 operate at the same period T, we can define a MIMO discrete-time LTI operator relating the sequences $(E^{m+}u)^T$ and $(E^{n+}y)^T$ appearing at the outputs of the two samplers. The impulse response of this operator is simply the sampled impulse response of $E^{n+}GE^{m-}$. We call this operator the Kranc operator for this multi-rate system:

$$G^T_{Kranc}(z) = (E^{n+}G(s)E^{m-})^T(z)$$

An equivalent representation for this system is shown in Figure 2.8-3. The continuous-time input u feeds into E^{m+}, is sampled at period T, and goes through the discrete-time Kranc operator $(E^{n+}GE^{m-})^T$. The discrete-time output of the Kranc operator passes through E^{n-} and emerges as a discrete-time sequence with period T/n. Note that there is still no transfer function describing this entire system, since the input is continuous and the output is discrete; however, the Kranc operator's transfer function $G^T_{Kranc}(z)$ contains all the information necessary to simulate the input-output behavior of this system for any arbitrary input.

If the Kranc operator is treated as a rational function in z, analysis can be difficult, since the Kranc operator is MIMO. It was shown in [22] that a state-space quadruple can be readily constructed for $G^T_{Kranc}(z)$, and that the dimension of the state space is no greater than that of the sample-equivalent system $G^{T/k}(z)$ sampled at the fast period T/k, where $k = \text{lcm}(m, n)$ is the least common multiple of m and n. If $G(s)$ has the form $G(s) = G_0(s)H_0^{T/m}(s)$ where $H_0^{T/m}(s)$ is a zero-order hold of length T/m and $G_0(s)$ is a rational function in s, then the Kranc operator's state dimension is equal to that of $G_0(s)$. This low state dimension makes the state-space formulation of Kranc operators a very efficient analysis technique for multi-rate systems. It is especially useful when a feedback system can be broken up into several multi-rate subsystems of the type in Figure 2.8-1. The Kranc operator for each subsystem can be computed separately, then the state-space quadruple for the entire closed loop system can be readily constructed using only real matrix arithmetic. The stability of the closed-loop system is tested by computing the eigenvalues of the A matrix for the closed-loop system. Section 2.2.1 describes the state-space operations required for common block diagram manipulations such as cascading and feedback interconnection.

The state-space realizations for Kranc operators, with and without a zero-order hold present, are given below.

Kranc operator without a ZOH. For any analog plant with input sampled at period T/m and output sampled at T/n, where m and n are integers, let

$$G^{T/k}(z) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

be the z-transform of the analog plant's impulse response sampled at period T/k, where $k = \mathrm{lcm}(m,n)$. The state-space quadruple of the resulting Kranc operator (sampled at period T) is

$$G^T_{Kranc}(z) = \left[\begin{array}{c|cccc} A^k & A^{k-1}B & A^{k-1-\frac{k}{m}}B & \cdots & A^{k-1-\frac{(m-1)k}{m}}B \\ \hline C & D_{11} & D_{12} & \cdots & D_{1m} \\ CA^{\frac{k}{n}} & D_{21} & \ddots & & \vdots \\ \vdots & \vdots & & & \vdots \\ CA^{\frac{(n-1)k}{n}} & D_{n1} & \cdots & \cdots & D_{nm} \end{array}\right]$$

where

$$D_{ij} = C\Psi\left(\frac{(i-1)k}{n} - \frac{(j-1)k}{m} - 1\right)B + \Omega_{ij}$$

$$\Psi(l) = \begin{cases} A^l & \text{if } l \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\Omega_{ij} = \begin{cases} D & \text{if } \frac{(i-1)k}{n} - \frac{(j-1)k}{m} = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Kranc operator with a ZOH. For an analog plant of the form $G(s) = G_0(s)H_0^{T/m}(s)$ (where $H_0^{T/m}(s)$ is a zero-order hold of length T/m), with input sampled at period T/m and output sampled at period T/n, compute the fast-rate sampled impulse response

$$(G_0 H_0^{T/k})^{T/k}(z) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

with a short zero-order hold of length T/k, where again $k = \mathrm{lcm}(m,n)$. Then the following formula gives a minimal realization for the Kranc operator:

$$G^T_{Kranc}(z) = (E^{n+} G_0 H_0^{T/m} E^{m-})^T$$

$$= \left[\begin{array}{c|cccc} A^k & A^{\frac{(m-1)k}{m}}XB & A^{\frac{(m-2)k}{m}}XB & \dots & XB \\ \hline C & D_{11} & D_{12} & \dots & D_{1m} \\ CA^{\frac{k}{n}} & D_{21} & \ddots & & \vdots \\ \vdots & \vdots & & & \vdots \\ CA^{\frac{(n-1)k}{n}} & D_{n1} & \dots & \dots & D_{nm} \end{array}\right]$$

where

$$X = \sum_{l=0}^{\frac{k}{m}-1} A^l$$

$$D_{ij} = C\left[\sum_{l=0}^{\frac{k}{m}-1} \Psi\left(l + \frac{(i-1)k}{n} - \frac{(j-1)k}{m} - \frac{k}{m}\right)\right] B + \Omega_{ij}$$

$$\Psi(l) = \begin{cases} A^l & \text{if } l \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\Omega_{ij} = \begin{cases} D & \text{if } 0 \leq \frac{(i-1)k}{n} - \frac{(j-1)k}{m} < \frac{k}{m} \\ 0 & \text{otherwise.} \end{cases}$$

2.8.2 Robustness Analysis Using Kranc Operators

Consider the multi-rate feedback system in Figure 2.8-4. Two analog plants $G(s)$ and $K(s)$ are connected in a feedback arrangement. Their outputs are sampled at periods T/n and T/m, where m and n are integers. Furthermore, the plant G is multiplied by an uncertain scalar quantity $1+\alpha$, where α is a constant but unknown complex number. If this system is known to be stable when $\alpha = 0$, a reasonable test of robustness is to find the largest real number β such that the system remains stable for all $|\alpha| < \beta$. The quantity β is sometimes called the gain-phase margin.

The Kranc equivalent system is shown in Figure 2.8-5. The Kranc operator for the upper plant is $(1+\alpha)G^T_{Kranc}$ where $G^T_{Kranc} = (E^{n+}GE^{m-})^T$, and for the lower plant it is $K^T_{Kranc} = (E^{m+}KE^{n-})^T$. Since $1+\alpha$ is constant at all frequencies, it may be pulled out of the Kranc operator and treated as a scalar multiplying the vector input to G^T_{Kranc}. Decomposition and recomposition operators E^{m+} and E^{n-} are required outside the feedback loop to vectorize the input r before sampling and to reconstruct the slow-rate vector output of G^T_{Kranc} into the fast-rate sequence $(y)^{T/n}$.

It is shown in [21] that, if this system is stable for $\alpha = 0$, it will remain stable for all $|\alpha| < \beta$ if and only if

$$\beta \leq \min_i \inf_\omega \left|1 + \lambda_i^{-1}\left(K^T_{Kranc}G^T_{Kranc}(e^{j\omega T})\right)\right|$$

where λ_i are the eigenvalues of $K^T_{Kranc}G^T_{Kranc}(e^{j\omega T})$. The closer the eigenvalues of the nominal loop gain come to the critical point -1, the smaller the robustness measure becomes. Computing this quantity requires finding the eigenvalues of a complex-valued matrix. The computation should be carried out for many frequency points; then the minimum value of the function can easily be seen on a plot of $\left|1 + \lambda_i^{-1}(K^T_{Kranc}G^T_{Kranc}(e^{j\omega T})\right|$ versus frequency. This technique will be used in Chapter 5 to determine the conservativeness of the new conic sector methods in the multirate example presented there.

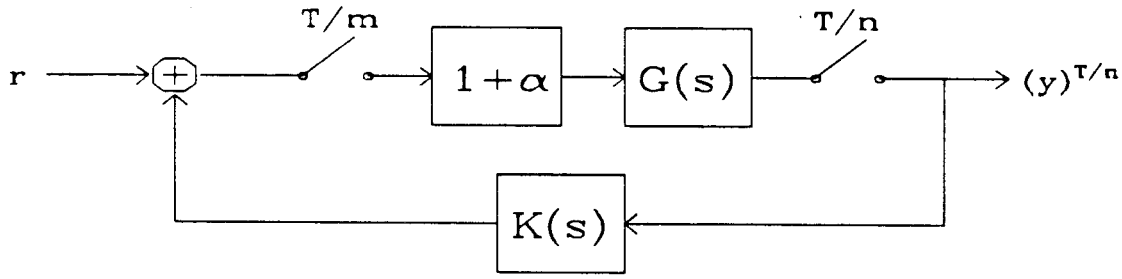

Fig. 2.8-4: A multirate feedback system with multiplicative perturbation $1 + \alpha$. α is assumed constant.

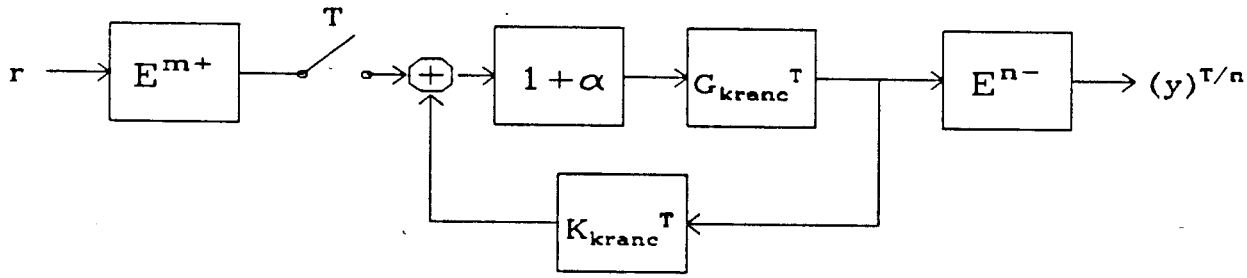

Fig. 2.8-5: Kranc operator equivalent of the above system. Feedback loop is now discrete-time LTI.

Analogous results for the gain margin and phase margin of a multirate system are also given in [21]. The eigenvalues λ_i, or characteristic loci, of the loop gain $K^T_{Kranc} G^T_{Kranc}(e^{j\omega T})$ are computed; then the gain margin or phase margin for each individual locus λ_i is found by the conventional SISO method. That is, the gain margin of the locus λ_i is the nonnegative real interval $[\alpha_{\min}(\lambda_i), \alpha_{\max}(\lambda_i)]$ where the locus λ_i can be scaled by any number in the interval without crossing the critical point -1. The phase margin is the maximum phase shift $\phi_{\max}(\lambda_i)$ the locus can be rotated by, in either direction, without crossing -1.

For the multirate system, the gain margin is the intersection of the gain margins lf the individual loci [21]:

$$GM = \bigcap_{i=1}^{m} [\alpha_{\min}(\lambda_i), \alpha_{\max}(\lambda_i)]$$

and the phase margin is the minimum of the phase margins of the individual loci:

$$PM = \min_i \phi_{\max}(\lambda_i)$$

Chapter Three — Conic Sectors for Hybrid Operators

3.1 Basic Theorem: a Cone that Contains the Hybrid Operator

This section presents a theorem from [20],[29] which is the basis for this thesis. The theorem allows us to construct a conic sector to contain any stable hybrid operator $\mathcal{K}$ of the form shown in Figure 3.1-1: a prefilter $F(s)$ followed by a sampler with period T, a digital filter $D(z)$, and a hold $H(s)$. Before presenting the theorem, let us examine in detail the input-output behavior of this hybrid operator.

This system is linear, though not time-invariant; therefore we can predict its output for any input in L_2 if we know its output for all complex exponential inputs $u_0 e^{j\omega t}$, where u_0 is any constant complex column vector. In the following derivation, let $\omega_s = 2\pi/T$.

Suppose the input is $u(t) = u_0 e^{j\omega T}$, where u_0 is a constant column vector $u_0 \in C^n$. Then the output of the sampler is

$$x_1 = \frac{1}{T}\sum_k F(j\omega)u_0 e^{j(\omega+k\omega_s)t}$$

The output of the digital filter is

$$x_2 = \frac{1}{T}\sum_k D(e^{j\omega T})F(j\omega)u_0 e^{j(\omega+k\omega_s)t}$$

Finally, the output of the hold element is

$$y = \frac{1}{T}\sum_k H(j\omega + jk\omega_s)D(e^{j\omega T})F(j\omega)e^{j(\omega+k\omega_s)t}$$

Now let $u(t)$ be any signal in L_2; u has a Fourier transform representation

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} U(j\omega)e^{j\omega t}d\omega$$

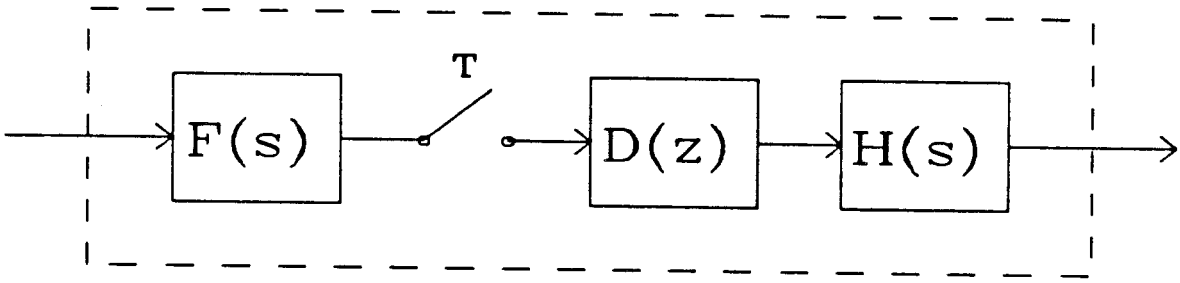

Fig. 3.1-1: A hybrid operator.

By invoking the linearity of the system, we can find the output y with this input u:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{T}\sum_k H(j\omega + jk\omega_s)D(e^{j\omega T})F(j\omega)U(j\omega)e^{j(\omega+k\omega_s)t}d\omega$$

By making a change of variables, this becomes

$$y(t) = \frac{1}{2\pi}\frac{1}{T}\int_{-\infty}^{\infty}\left[\sum_k H(j\omega)D(e^{j\omega T})F(j\omega - jk\omega_s)U(j\omega - jk\omega_s)\right]e^{j\omega t}d\omega$$

where we have taken advantage of the periodicity of $D(e^{j\omega T})$. Given this input-output description of the system, we may proceed to derive a conic sector which strictly contains it.

Let the cone center C be any LTI L_{2e}-stable operator; we will see shortly that the optimal choice for C is given by the baseband (non-aliased) response of the operator K:

$$C_{\text{opt}} = \frac{1}{T}HDF.$$

In order to simplify the notation, define

$$K_n(j\omega) = \begin{cases} \frac{1}{T}H(j\omega)D(e^{j\omega T})F(j\omega) - C(j\omega) & \text{if } n = 0; \\ \frac{1}{T}H(j\omega)D(e^{j\omega T})F(j\omega - jn\omega_s) & \text{otherwise.} \end{cases}$$

Note that for an input $u(t)$ to this system,

$$y(t) - (Cu)(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_n K_n(j\omega)U(j\omega - jn\omega_s)e^{j\omega t}d\omega$$

In order to assure Lebesgue Dominated Convergence, which is required in one step of the derivation, assume for $|\omega|$ sufficiently large that

$$\sum_n \|K_n(j\omega)\|^2 \leq \frac{\alpha}{|\omega|^{1+\beta}} \quad \text{for some } \alpha,\ \beta > 0.$$

Now we may derive the following Lemma, which is the same as Lemma 3.A of [20].

Lemma. Let $e(s)$ be the Laplace transform of $e(t) \in L_2$, and let K_n be as described above. Then

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\sum_n K_n(j\omega)e(j\omega - jn\omega_s)\right\|^2 d\omega \\ \leq\ &\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_k\sum_n \|K_n(j\omega + jk\omega_s)\|^2\right)\|e(j\omega)\|^2 d\omega \end{aligned}$$

Proof:

$$\begin{aligned}
& \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\sum_n K_n(j\omega)e(j\omega - jn\omega_s)\right\|^2 d\omega \\
\leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_n \|K_n(j\omega)e(j\omega - jn\omega_s)\|\right)^2 d\omega \\
\leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_n \|K_n(j\omega)\|\,\|e(j\omega - jn\omega_s)\|\right)^2 d\omega \\
\leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_n \|K_n(j\omega)\|^2\right)\left(\sum_k \|e(j\omega - jk\omega_s)\|^2\right) d\omega
\end{aligned}$$

Now move one of the summation signs outside the integral by Lebesgue Dominated Convergence:

$$= \frac{1}{2\pi}\sum_k \int_{-\infty}^{\infty}\left(\sum_n \|K_n(j\omega)\|^2\right)\|e(j\omega - jk\omega_s)\|^2 d\omega$$

Next make a change of variables:

$$\begin{aligned}
&= \frac{1}{2\pi}\sum_k \int_{-\infty}^{\infty}\left(\sum_n \|K_n(j\omega + jk\omega_s)\|^2\right)\|e(j\omega)\|^2 d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_k\sum_n \|K_n(j\omega + jk\omega_s)\|^2\right)\|e(j\omega)\|^2 d\omega
\end{aligned}$$

This completes the proof of the Lemma.

Now we may proceed to prove the main theorem. For simplicity of notation, define $H_k(j\omega) = H(j\omega - jk\omega_s)$. Define $F_k(j\omega)$ and $C_k(j\omega)$ similarly.

Theorem 1. Let K be a stable hybrid operator, let C be any LTI L_{2e}-stable operator, and let R and R^{-1} be LTI L_{2e}-stable operators.

(a) K is strictly inside Cone(C, R) if

$$\underline{\sigma}^2(R(j\omega)) \geq \frac{1}{1-\epsilon}(r_1(\omega) - r_2(\omega) + r_3(\omega)) \qquad \forall\, \omega \in R$$

for some $1 > \epsilon > 0$, where

$$
\begin{aligned}
r_1(\omega) &= \frac{1}{T^2}\left(\sum_k \bar{\sigma}^2[H_k(j\omega)]\right)\bar{\sigma}^2[D(e^{j\omega T})]\left(\sum_n \bar{\sigma}^2[F_n(j\omega)]\right) \\
r_2(\omega) &= \frac{1}{T^2}\sum_k \bar{\sigma}^2[H_k(j\omega)D(e^{j\omega T})F_k(j\omega)] \\
r_3(\omega) &= \sum_k \bar{\sigma}^2\left(\frac{1}{T}H_k(j\omega)D(e^{j\omega T})F_k(j\omega) - C_k(j\omega)\right)
\end{aligned}
$$

(b) Furthermore, the optimal center

$$
C(j\omega) = \frac{1}{T}H(j\omega)D(e^{j\omega T})F(j\omega)
$$

minimizes the lower bound for $\underline{\sigma}(R(j\omega))$.

Proof: define the truncated function

$$
\tilde{e}_\tau(t) = \begin{cases} (Re)(t) & \text{if } t \geq \tau; \\ 0 & \text{if } t < \tau. \end{cases}
$$

then

$$
\begin{aligned}
& \|(K-C)e\|_\tau^2 \\
= \; & \|(K-C)R^{-1}Re\|_\tau^2 \\
= \; & \|(K-C)R^{-1}\tilde{e}_\tau\|_\tau^2 \\
\leq \; & \|(K-C)R^{-1}\tilde{e}_\tau\|_{L_2}^2 \\
= \; & \int_0^\infty \left\|(K-C)R^{-1}\tilde{e}_\tau(t)\right\|^2 dt \\
= \; & \frac{1}{2\pi}\int_{-\infty}^\infty \left\|(K-C)R^{-1}\tilde{e}_\tau(j\omega)\right\|^2 d\omega \quad \text{(by Parseval's theorem)} \\
= \; & \frac{1}{2\pi}\int_{-\infty}^\infty \left\|\sum_n K_n(j\omega)R^{-1}(j\omega - jn\omega_s)\tilde{e}_\tau(j\omega - jn\omega_s)\right\|^2 d\omega
\end{aligned}
$$

Now applying the above Lemma:

$$\begin{aligned}
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_k\sum_n \|K_n(j\omega+jk\omega_s)\|^2\right)\left\|R^{-1}(j\omega)\tilde{e}_\tau(j\omega)\right\|^2 d\omega \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(r_1(\omega)-r_2(\omega)+r_3(\omega)\right)\left\|R^{-1}(j\omega)\tilde{e}_\tau(j\omega)\right\|^2 d\omega \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}(1-\epsilon)\underline{\sigma}^2(R(j\omega))\left\|R^{-1}(j\omega)\right\|^2\|\tilde{e}_\tau(j\omega)\|^2 d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}(1-\epsilon)\|\tilde{e}_\tau(j\omega)\|^2 d\omega \\
&\leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\|\tilde{e}_\tau(j\omega)\|^2 d\omega - \epsilon\frac{1}{2\pi}\int_{-\infty}^{\infty}\|R(j\omega)\|^2\left\|R^{-1}\tilde{e}_\tau(j\omega)\right\|^2 d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\|\tilde{e}_\tau(j\omega)\|^2 d\omega - \epsilon'\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\|R(j\omega)\|^2}{\inf\limits_{\omega\in R}\|R(j\omega)\|^2}\left\|R^{-1}\tilde{e}_\tau(j\omega)\right\|^2 d\omega
\end{aligned}$$

where $\epsilon' = \epsilon\left(\inf\limits_{\omega\in R}\|R(j\omega)\|^2\right) > 0$.

$$\begin{aligned}
&\leq \|\tilde{e}_\tau\|_{L_2}^2 - \epsilon'\|R^{-1}\tilde{e}_\tau\|_{L_2}^2 \\
&\leq \|\tilde{e}_\tau\|_\tau^2 - \epsilon'\|R^{-1}\tilde{e}_\tau\|_\tau^2
\end{aligned}$$

and by causality of R^{-1}:

$$\begin{aligned}
&= \|Re\|_\tau^2 - \epsilon'\|R^{-1}Re\|_\tau^2 \\
&= \|Re\|_\tau^2 - \epsilon'\|e\|_\tau^2 \\
&\leq \|Re\|_\tau^2 - \epsilon''\left(\|e\|_\tau^2 + \|y\|_\tau^2\right)
\end{aligned}$$

where

$$\epsilon'' = \frac{\epsilon'}{1+\alpha^2} > 0 \quad \text{and} \quad \alpha = \sup_{e,\tau}\left(\frac{\|Ke\|_\tau}{\|e\|_\tau}\right) < \infty \quad \text{since } K \text{ is stable.}$$

So the system K meets the definition of being strictly inside $\text{Cone}(C,R)$. This completes the proof of part (a). The proof of part (b) is simple: observe that $r_3(\omega) = 0$ when $C = C_{\text{opt}} = \frac{1}{T}HDF$. Since $r_1(\omega)$ and $r_2(\omega)$ are independent of C, this choice for C minimizes $r_1(\omega) - r_2(\omega) + r_3(\omega)$. This completes the proof of Theorem 1.

In the SISO case, the terms r_1 and r_2 may be combined in closed form to yield a simpler formula:

Theorem 1'. Let K be a stable SISO hybrid operator, let C be the optimal (stable) cone center $C(j\omega) = \frac{1}{T}H(j\omega)D(e^{j\omega T})F(j\omega)$, and let R and R^{-1} be LTI L_{2e}-stable operators. Then K is stricly inside Cone(C, R) if

$$|R(j\omega)|^2 \geq \frac{1}{1-\epsilon}\left(\frac{1}{T^2}\sum_k \sum_{n\neq k} |H_k(j\omega)F_n(j\omega)|^2\right)|D(e^{j\omega T})|^2$$

for some $\epsilon > 0$.

This version of the theorem will be used extensively in this thesis. Chapter 4 presents a method from [20] for computing this lower bound for $|R|$ in closed form, when H and F are SISO rational functions in s.

3.2 Limitations of the Basic Theorem

The preceding theorem can be used directly to determine the stability of a nominal sampled-data feedback system as long as the hybrid compensator is stable and certain other assumptions are met. Consider the system in Figure 3.2-1. If all the system components $G(s)$, $H(s)$, $D(z)$, and $F(s)$ are exactly known, we can use Theorem 1 and the results of Section 2.4 to check a sufficient condition for closed-loop stability as follows:

1. Use Theorem 1 to compute $C(j\omega)$, $R(j\omega)$ so that the hybrid compensator K is strictly inside Cone(C, R).

2. Using a Nyquist plot or some other method, test for stability of the system when $C(j\omega)$ is substituted for the hybrid compensator K.

3. Compute $\bar{\sigma}(RG(I+CG)^{-1})$ at all frequencies. if C satisfies the test in (2) and $\|RG(I+CG)^{-1}\|_\infty < 1$, then the conic sector stability criterion is satisfied and the system is stable.

It is important to note, however, that this is only a sufficient condition. It is easy to construct examples which are closed-loop stable but which have $\|RG(I-CG)^{-1}\|_\infty > 1$. For SISO systems, this condition is caused solely by slow rolloff at high frequencies in F and H. An examination of the formula for $R(j\omega)$ in Theorem 1 reveals that, when the optimal cone center $\frac{1}{T}HDF$ is used, the cone radius consists entirely of cross-products $\bar{\sigma}^2(H_k)\bar{\sigma}^2(D)\bar{\sigma}^2(F_n)$ where $k \neq n$. That is, every term of $R^2(j\omega)$ involves the product of the gains of F and H at different aliased frequencies $\omega + k\omega_s$ and $\omega + n\omega_s$. If the prefilter F and the hold H have very low gain at frequencies above the Nyquist rate π/T, all of these cross-products will be nearly zero and the radius will be very small. In the limit as perfect prefiltering is achieved, the radius R goes to zero. This can be interpreted as follows: with perfect prefiltering ($F(j\omega) = 0$ and $H(j\omega) = 0$ for $\omega > \pi/T$), no aliasing or "stairstepping"

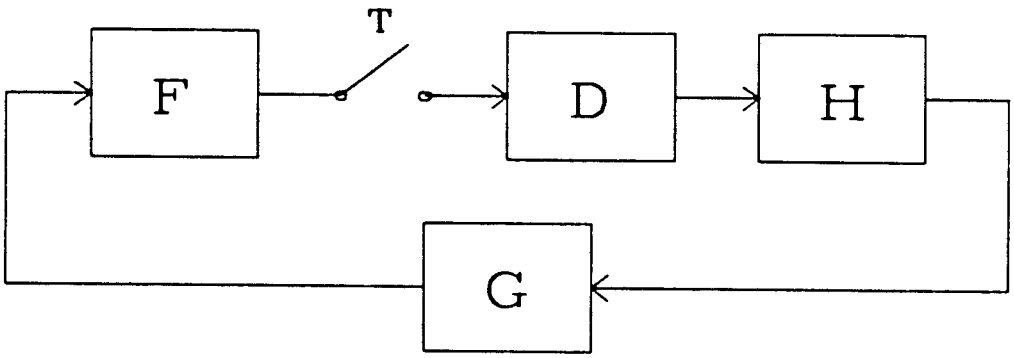

Fig. 3.2-1: A sampled-data feedback loop.

occurs, and the hybrid operator no longer has a time-varying response. It becomes an LTI system completely described by the LTI cone center C.

Another source of conservativeness occurs only in the MIMO case. Since the formula for R involves maximum singular values, considerable information about the system is lost when $D(z)$ is poorly conditioned; i.e., when $\kappa = \bar{\sigma}(D)/\underline{\sigma}(D) \gg 1$ at some frequencies. This causes stable systems to fail the sufficient condition for stability when the controller has a high condition number κ. An example of this is given in Chapter 5.

An obvious limitation of Theorem 1 is that is requires the hybrid compensator to be stable. This rules out the direct use of the theorem for integrating compensators such as PI and PID controllers, as well as many compensators for open-loop unstable systems.

One more limitation in the MIMO case is that all channels must be sampled at the same period T. This rules out direct application of the theorem to multi-rate systems.

3.3 Flexibility of the Basic Theorem

All the limitations of Theorem 1 mentioned above can be overcome by exploiting the inherent flexibility of the theorem. Although it is natural to think of the operators F and H as corresponding to the actual anti-aliasing filter and hold element that are physically present in the system, F and H can be set equal to any arbitrary transfer functions, so long as they are stable. Refer to Figure 3.3-1; a natural application of Theorem 1 would have us finding a cone bound for the hybrid compensator in the dashed box — F, the sampler, D, and H. However, we can choose any stable, invertible transfer functions $\alpha(s)$ and $\beta(s)$ with stable inverses and restructure the system as shown in Figure 3.3-2. α and β appear in series

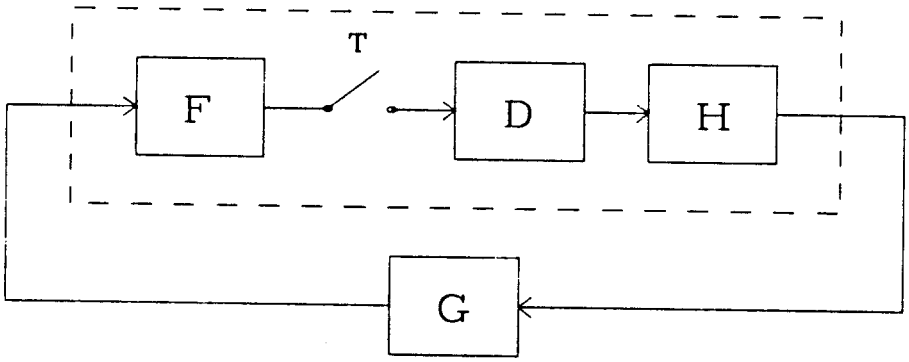

Fig. 3.3-1: A sampled-data feedback loop. The subsystem in the dashed box can be cone-bounded directly by Theorem 1.

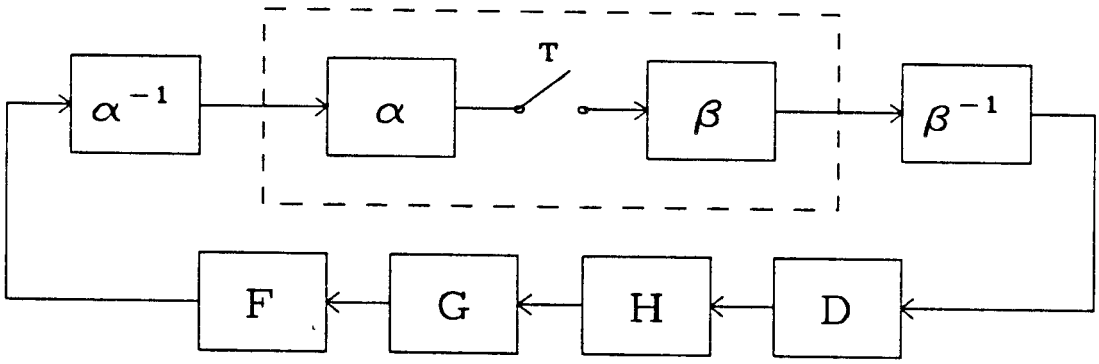

Fig. 3.3-2: By inserting fictitious operators α, β, α^{-1}, and β^{-1}, the conservativeness of Theorem 1 can be reduced.

with their inverses, so they have no effect on the closed-loop system. By applying Theorem 1 to the fictitious hybrid compensator in the dashed box — α, the sampler, and β — and lumping F, α^{-1}, β^{-1}, D, and H with the plant G, we get an alternative and equally valid sufficient condition for closed-loop stability. Since the only conditions on α and β are that they be stable with stable inverses, we actually get a family of tests for closed-loop stability.

Let K denote the hybrid operator consisting of α, the sampler, and β. Then Theorem 1 results in C, R such that K is inside Cone(C, R) where

$$C(j\omega) = \frac{1}{T}\beta(j\omega)\alpha(j\omega)$$

$$\underline{\sigma}^2(R(j\omega)) = \frac{1}{T^2}\left\{\left(\sum_k \bar{\sigma}(\beta_k(j\omega))\right)\left(\sum_k \bar{\sigma}(\alpha_n(j\omega))\right) - \sum_k \bar{\sigma}(\beta_k(j\omega)\alpha_k(j\omega))\right\}$$

For simplicity in the MIMO case, assume R to be a diagonal operator with the squared magnitude of each diagonal element $R_{ii}(j\omega)$ equal to the expression above. Further assume that R_{ii} is minimum phase; this satisfies the requirement that R and R^{-1} both be L_{2e}-stable. Denote the cone radius so defined by $R = \text{Rad}(\alpha, \beta)$. Such a minimum-phase factor always exists: Chapter 6 presents a method for approximating it with a rational function. However, since R is a scalar times an identity matrix, and we already know the magnitude of the scalar, it is not necessary to know the phase of R_{ii} at each frequency in order to test the closed-loop stability condition.

For this system, the closed-loop stability criterion is

$$\|R\alpha^{-1}FGHD\beta^{-1}(I - C\alpha^{-1}FGHD\beta^{-1})^{-1}\|_\infty < 1$$

where R and C are defined as above. This expression will be simplified somewhat in Section 3.4. For now, simply note that the above expression provides a whole family of sufficient tests for stability, parameterized by the operators α and β. The following sections will show how, by making appropriate choices for α and β, to

overcome all the limitations of Theorem 1 pointed out in the preceding section. One improvement is already obvious: the requirement that $D(z)$ be stable is no longer needed. Since D has been pulled outside the hybrid operator K and lumped with the plant, the restriction of Theorem 1 that K be stable has no influence on D.

3.3.1 A Simple Cone for a Sampler with a ZOH

Consider the SISO hybrid operator in Figure 3.3-3. It consists of a sampler sandwiched between two identical operators $\alpha = 1/(s+\epsilon)$, where $\epsilon > 0$ is vanishingly small. We cannot simply use $1/s$, because then the operator would not be L_{2e}-stable and Theorem 1 would not apply. This operator is inside Cone(C, R) where $C = \frac{1}{T}\alpha^2$ and R =Rad(α, α) as defined above. An examination of the formula defining R reveals that $R(j\omega)$ is periodic in ω with period $2\pi/T$; therefore, it can be considered to be the transfer function of a discrete-time operator evaluated on the unit circle: $R(z)$ where $z = e^{j\omega T}$. [20] presents a method to evaluate $|\text{Rad}(\alpha, \beta)|^2$ in closed form as a rational function in z when α and β are SISO rational functions of finite order. The method is described in Section 4.1.1 of this thesis. Application of this method to the hybrid operator in Figure 3.3-3 gives this result:

$$|R(j\omega)|^2 = \frac{-W(\epsilon T)z^{-1}(z^{-2} - (e^{-\epsilon T} + e^{\epsilon T})z^{-1} + 1)}{4\epsilon^3 T(1 - e^{-\epsilon T}z^{-1})^2(1 - e^{\epsilon T}z^{-1})^2}$$

where $z = e^{j\omega T}$ and

$$W(\epsilon T) = (\epsilon T + 1)e^{-\epsilon T} + (\epsilon T - 1)e^{\epsilon T}$$

Since for $\epsilon T \ll 1$, $W(\epsilon T) \approx \frac{2}{3}\epsilon^3 T^3$, this expression simplifies greatly in the limit:

$$\lim_{\epsilon \to 0} |R(j\omega)|^2 = \frac{-T^2 z^{-1}}{6(1 - z^{-1})^2}$$

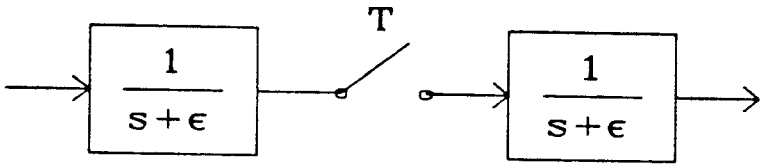

Fig. 3.3-3: A sampler sandwiched between two near-integrators $\alpha = 1/(s+\epsilon)$.

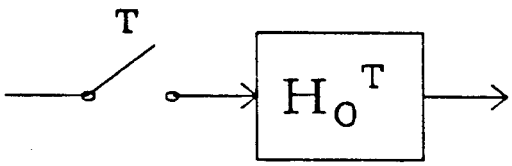

Fig. 3.3-4: A sampler followed by a ZOH.

In spite of the minus sign, this expression is positive everywhere on the unit circle.

Now consider the system in Figure 3.3-4: a sampler followed by a ZOH. The system is redrawn in Figure 3.3-5 with α and α^{-1} inserted in series before and after the sampler. The result above gives us a cone bound for the operator in the dashed box; we would like to extend this to a cone bound for the entire system. In Section 2.3, it was shown that if an operator K in Cone(C, R) is preceded by an LTI operator F, the cascaded operator KF is inside Cone(CF, RF). We can use this property to lump the α^{-1} operator at the input of the dashed box inside a cone bound, but what can be done about the operators α^{-1} and H_0^T at the output of the box? Note that, since $H_0^T(j\omega) = (1 - e^{-j\omega T})/j\omega$, the ZOH has a factor which is periodic in ω with period $2\pi/T$. Since the operator in the box is PLTV with period T, the results of Section 2.7.4 allow us to commute the box with the periodic factor of the ZOH. This leads to Figure 3.3-6, in which the box is preceded by $(s+\epsilon)(1 - e^{-sT})$ and followed by $(s+\epsilon)/s$. But $\lim_{\epsilon \to 0}(s+\epsilon)/s = 1$; therefore in the limit we can ignore the operator at the output of the box and use the results of 2.3 to produce a cone bound for the entire system: the sampler followed by a ZOH is inside Cone(C_0, R_0) where

$$\begin{aligned} C_0(j\omega) &= \lim_{\epsilon \to 0} \frac{1}{T}\alpha^2\left(\frac{1 - z^{-1}}{\alpha}\right) = \frac{1}{T}\frac{1 - z^{-1}}{s} \\ &= \frac{1}{T}H_0^T(s) \\ |R_0(j\omega)|^2 &= \lim_{\epsilon \to 0} |\mathrm{Rad}(\alpha, \alpha)|^2 |(s+\epsilon)(1 - z^{-1})|^2 \\ &= \frac{-T^2 z^{-1}}{6(1 - z^{-1})^2}(-s^2(1 - z^{-1})(1 - z)) \\ &= \frac{T^2 s^2}{6} \end{aligned}$$

and $z = e^{j\omega T}$, $s = j\omega$ as usual.

This is a remarkably simple formula, and its behavior at low frequencies agrees well with intuition. Since the output of this system is a stairstep waveform which

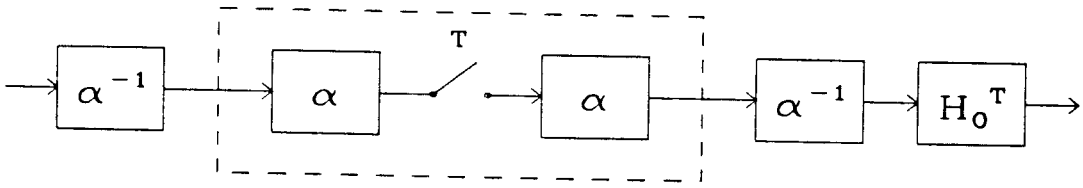

Fig. 3.3-5: A sampler and ZOH, with $\alpha = 1/(s+\epsilon)$ and α^{-1} inserted. The subsystem in the dashed box has squared cone radius $R^2 = -T^2 e^{j\omega T}/(6(1-e^{-j\omega T})^2)$.

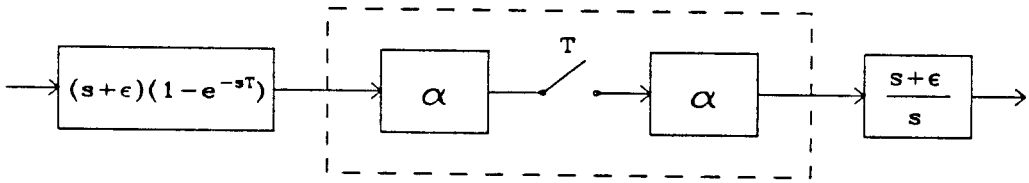

Fig. 3.3-6: The same system with the $(1-e^{-sT})$ factor of H_0^T commuted to the input. This is legal since the system is PLTV.

equals the input at sampling instants $t = kT$, the output will closely approximate the input when the input is slowly varying. This means that for a low-frequency sine wave input $u = A\sin(\omega t)$, the output error $y - Cu$ goes to zero as $\omega \to 0$. The radius R_0 given above is also zero at DC, so it correctly predicts this property of the system. Let's examine this more closely.

For low-frequency sine wave inputs, the cone center $C = \frac{1}{T}H_0^T$ is closely approximated by a time delay of length $T/2$: $C(j\omega) \approx e^{-j\omega T/2}$ for $\omega T \ll 1$. Figure 3.3-7 shows the output of the system and the output of C; the two waveforms intersect at the midpoint of each sampling interval. For ωT small, Cu is well approximated inside each sampling interval $t \in [kT, kT + T]$ by a straight line with slope $m = \frac{d}{dt}u(t) = A\omega\cos(\omega kT)$. We can integrate the square error over one sampling interval:

$$\int_{kT}^{kT+T} (y(t) - (Cu)(t))^2 dt \approx 2\int_0^{T/2} (mt)^2 dt = \frac{m^2T^3}{12}$$

If we integrate over a long period of time $\tau = NT$:

$$\|y - Cu\|_\tau^2 \approx \sum_{k=0}^{N} \frac{A^2\omega^2\cos^2(\omega kT)T^3}{12} \approx \frac{NA^2\omega^2T^3}{24}$$

Now compare this to the energy of Ru, where $R(s) = sT/\sqrt{6}$:

$$\|Ru\|_\tau^2 = \int_0^{NT} \left(\frac{A\omega T\sin(\omega T)}{\sqrt{6}}\right)^2 dt \approx \frac{NA^2\omega^2T^3}{12}$$

so we see that this approximation for $y - Cu$ satisfies the conic sector inequality $\|y - Cu\|_\tau < \|Ru\|_\tau$ for large $\tau > 0$. In fact, $\|y - Cu\|_\tau / \|Ru\|_\tau \approx 1/\sqrt{2}$; this means that for low-frequency sine wave inputs, the cone radius consistently overestimates the error by a factor of $\sqrt{2}$.

The radius R goes to infinity as $\omega \to \infty$. This behavior is necessary, since there is no prefilter at all in front of the sampler. The system's response to inputs with

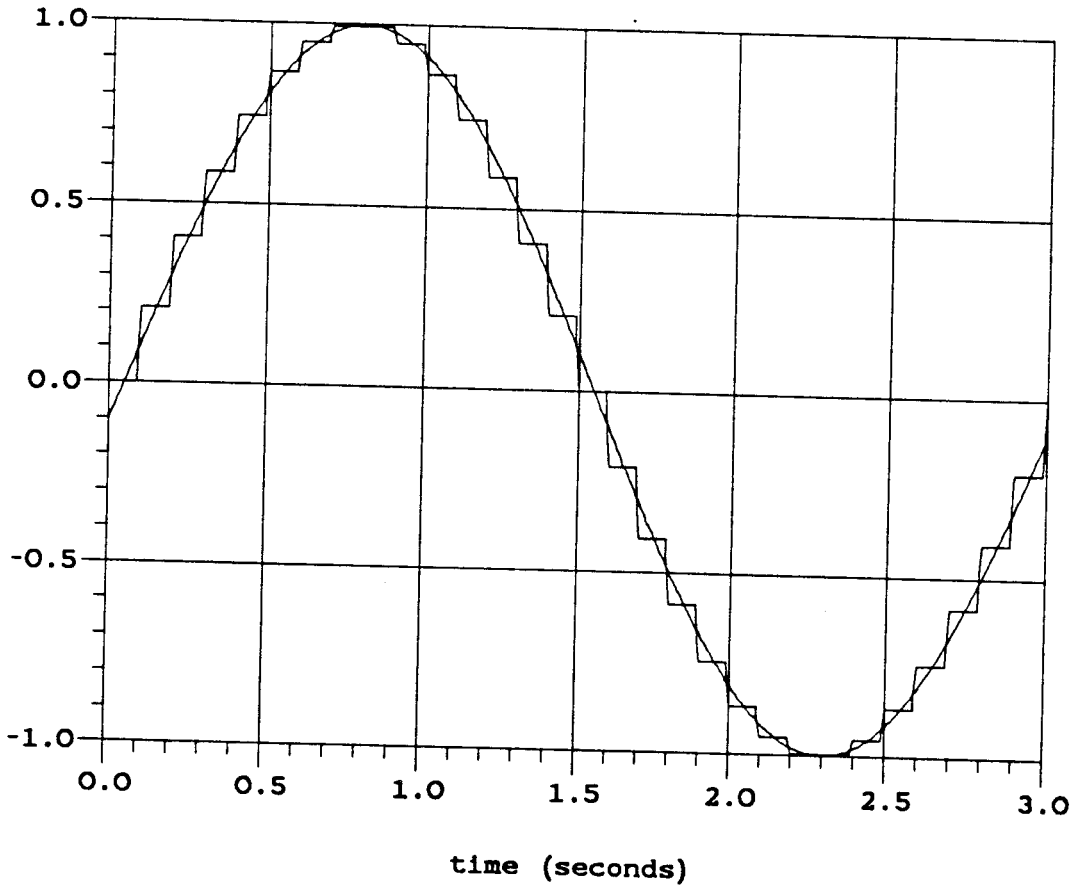

Fig. 3.3-7: The outputs of $C = \frac{1}{T} H_0^T$ and of the sampler plus ZOH, when the input is $u = \sin \omega t$ with $\omega = 2.094$ and $T = 0.1$ sec. Cu approximately bisects the ZOH output in each sampling interval.

energy at very high frequencies, e.g., white noise, is unbounded because the high-frequency components are aliased into the baseband, where they are not attenuated by the ZOH.

This conic sector for a sampler with hold is appealing because of its simplicity. Any SISO sampled-data feedback system with ZOH can be rearranged to take advantage of this simple formula. See the system in Figure 3.3-8. Since $D(z)$ and $H_0^T(s)$ are commutative, the hold can be placed immediately after the sampler, and F and D can be lumped with the plant G as in Figure 3.3-9. The conic sector stability criterion is now simply stated as

$$\|RFGD(1 - \frac{1}{T}H_0^T FGD)^{-1}\|_\infty \; < \; 1$$

where $R(s) = sT/\sqrt{6}$. Since R is independent of F and D, there is no need to recompute the infinite summations of Theorem 1 for each change in the system.

When the sampling rate $2\pi/T$ is much greater than the loop bandwidth of the system, this approach can give better results than a direct application of Theorem 1. In general, however, it is often still too conservative. The next section presents an optimal solution to the SISO analysis problem.

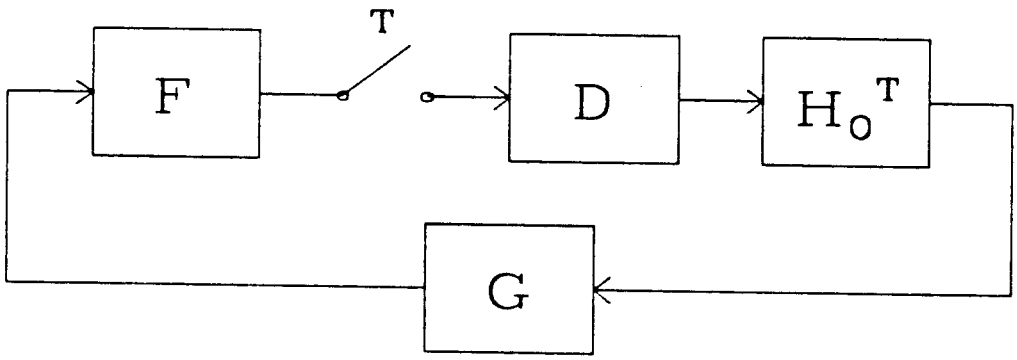

Fig. 3.3-8: A sampled-data feedback loop with a ZOH.

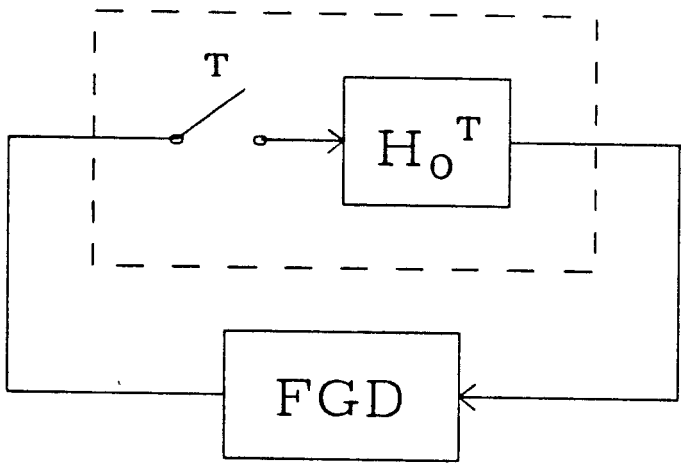

Fig. 3.3-9: The same system with D and H_0^T commuted, and with F and D lumped with G. The operator in the dashed box has cone radius $R = \omega T/\sqrt{6}$.

3.4 Optimal Cone Radius in SISO Case

Figure 3.4-1 shows a SISO hybrid feedback system with the fictitious operators α, β, and their inverses embedded around the sampler. We have already seen that a sufficient condition for stability of this system is

$$\|R\alpha^{-1}FGHD\beta^{-1}(I + C\alpha^{-1}FGHD\beta^{-1})^{-1}\|_\infty < 1$$

where $R = \text{Rad}(\alpha, \beta)$ and $C = \frac{1}{T}\beta\alpha$. This can be simplified in a way that is also valid for MIMO systems. As in the previous section, let K denote the hybrid operator consisting of α, the sampler, and β. Theorem 1 guarantees that K is inside Cone(C, R). We can split K into an LTI part C and a cone-bounded PLTV part with cone center zero: $K = C + \hat{K}$ where $\hat{K}$ is inside Cone$(0, R)$. Now redraw the system as in Figure 3.4-2, with K split into two parts and with α^{-1} and β^{-1} duplicated. Since $C = \frac{1}{T}\beta\alpha$, the path containing C simplifies to $\frac{1}{T}$ as in Figure 3.4-3. Finally, in Figure 3.4-4, the entire LTI portion of the remaining system is lumped into a single operator:

$$M = FGHD(I + \frac{1}{T}FGHD)^{-1}$$

The stability condition now is simply

$$\left\| \frac{RM}{\alpha\beta} \right\|_\infty < 1$$

Note that M is independent of α and β. This formulation suggests an optimization problem:

$$\inf_{\alpha,\beta} \left\| \frac{RM}{\alpha\beta} \right\|_\infty$$

The solution to this optimization problem provides us with the optimal, i.e., least conservative, sufficient condition for stability of a SISO system when Theorem 1 is used. Surprisingly, the optimal solution is trivially simple to compute. The solution and a proof of its optimality are presented below.

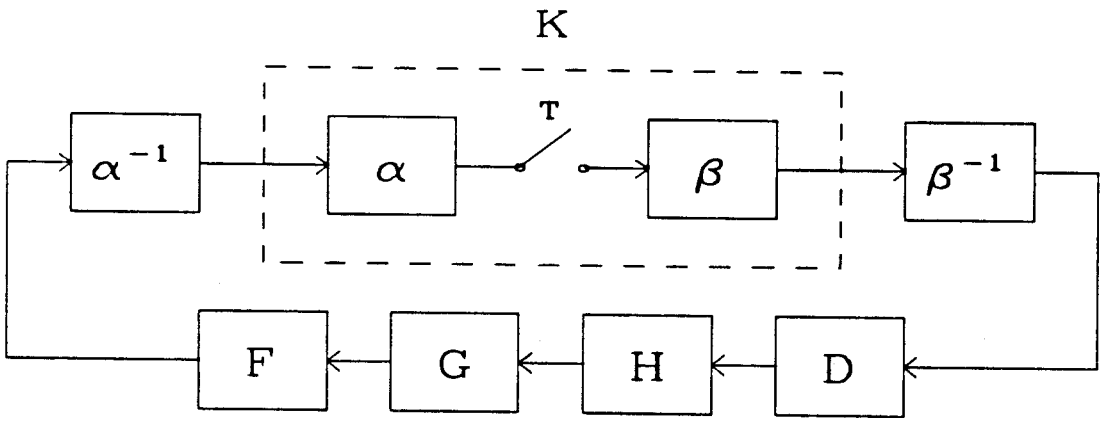

Fig. 3.4-1: A SISO hybrid feedback system with fictitious operators α, α^{-1}, β, β^{-1}. Call the PLTV operator in the dashed box K.

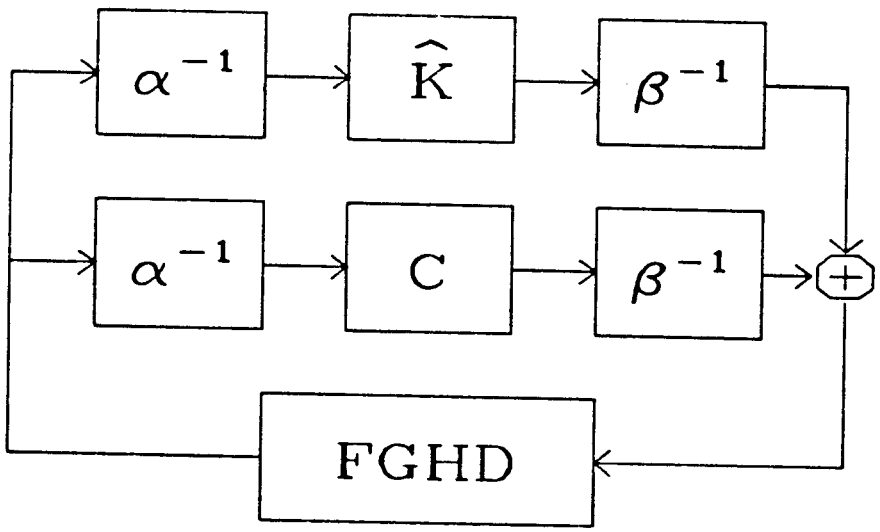

Fig. 3.4-2: Split K into an LTI part $C = \alpha\beta/T$ and a zero-center PLTV part $\hat{K}$ inside Cone$(0, R)$.

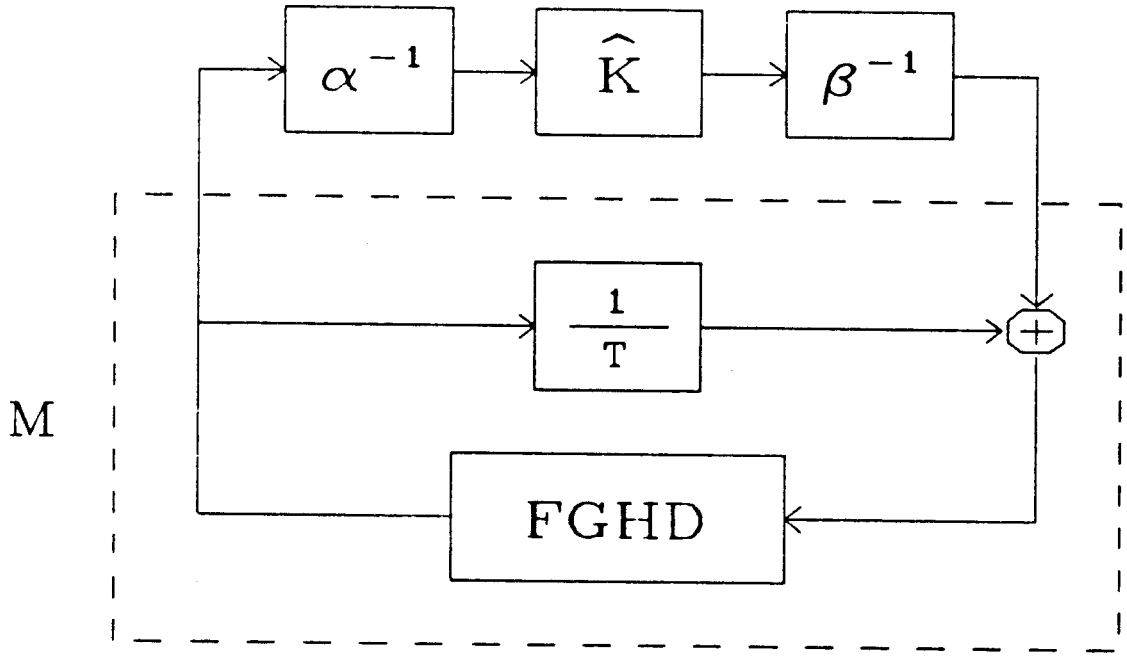

Fig. 3.4-3: Lump the lower path with $FGHD$ and call the result M.

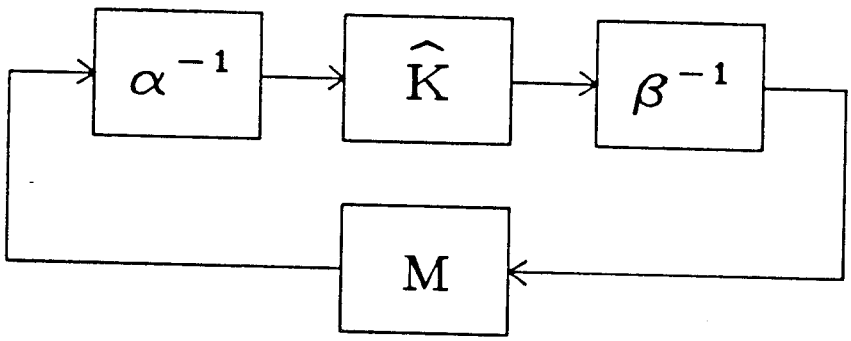

Fig. 3.4-4: The simplified system. M is stable if the original system is stable when the sampler is replaced by $1/T$.

3.4.1 Theorem and Proof

Since $\|RM/\alpha\beta\|_\infty$ depends only on the magnitudes of R, M, α, and β at each frequency, we will assume without loss of generality that each of these four functions is a real-valued function of a real variable. This will simplify the notation in the following theorem. For example, define $M(\omega)$ to equal $|M_0(j\omega)|$ for all $\omega \in R$, where M_0 is the original complex-valued function. Define $\alpha(\omega)$, $\beta(\omega)$, and $R(\omega)$ similarly.

To simplify the notation further, define

$$M_k(\omega) \equiv M(\omega + k\omega_s)$$

where $\omega_s = 2\pi/T$. Define $\alpha_k(\omega)$ and $\beta_k(\omega)$ similarly. This notation is not needed for R since it is periodic in ω with period ω_s.

With these definitions, the radius R is given by

$$\begin{aligned} R^2(\omega) &\equiv \frac{1}{T^2}\left[\left(\sum_k \alpha_k^2(\omega)\right)\left(\sum_k \beta_k^2(\omega)\right) - \left(\sum_k \alpha_k^2(\omega)\beta_k^2(\omega)\right)\right] \\ &= \frac{1}{T^2}\sum_k \sum_{n\neq k} \alpha_k^2(\omega)\beta_n^2(\omega) \end{aligned}$$

Now observe that since the radius $R(\omega)$ depends only upon the values of α and β at the aliased frequencies $\omega + k\omega_s$, the minimization problem

$$\inf_{\alpha,\beta} \left\| \frac{RM}{\alpha\beta} \right\|_\infty$$

is equivalent to a *family* of uncoupled minimization problems with ω fixed at a point on the interval $[0, \omega_s)$. In other words, we can fix ω at some point on this interval and choose $\alpha_k(\omega)$ and $\beta_k(\omega)$ for each integer k to minimize

$$\sup_k \left| \frac{R(\omega)M_k(\omega)}{\alpha_k(\omega)\beta_k(\omega)} \right|$$

If this is done for all $\omega \in [0, \omega_s)$, $\alpha(\omega)$ and $\beta(\omega)$ will be defined over their entire domain $\omega \in R$. Furthermore, since these choices for α and β will, by definition,

minimize

$$\sup_{k,\omega} \left| \frac{R(\omega)M_k(\omega)}{\alpha_k(\omega)\beta_k(\omega)} \right| = \left\| \frac{RM}{\alpha\beta} \right\|_\infty$$

this is a solution to the complete minimization problem.

Theorem 2. For the minimization problem with ω fixed:

$$\inf_{\alpha_k(\omega),\beta_k(\omega)} \sup_k \left| \frac{R(\omega)M_k(\omega)}{\alpha_k(\omega)\beta_k(\omega)} \right|$$

all optimal solutions are of the form

$$\alpha_k(\omega) = K_1\sqrt{M_k(\omega)}$$

$$\beta_k(\omega) = K_2\sqrt{M_k(\omega)}$$

where K_1 and K_2 are arbitrary real positive constants.

Proof: Define

$$W_k(\omega) \equiv \frac{R(\omega)M_k(\omega)}{\alpha_k(\omega)\beta_k(\omega)}$$

By definition,

$$R^2(\omega) = \frac{1}{T^2}\sum_k \sum_{n\neq k} \alpha_k^2(\omega)\beta_n^2(\omega)$$

so its derivatives with respect to $\alpha_i^2(\omega)$ and $\beta_i^2(\omega)$ are given by

$$\frac{d(R^2(\omega))}{d(\alpha_i^2(\omega))} = \frac{1}{T^2}\sum_{k\neq i} \beta_k^2(\omega)$$

$$\frac{d(R^2(\omega))}{d(\beta_i^2(\omega))} = \frac{1}{T^2}\sum_{k\neq i} \alpha_k^2(\omega)$$

Note that both derivatives are always positive. Now define $\|W_k(\omega)\|_\infty \equiv \sup_k W_k(\omega)$ with ω fixed. Suppose $W_i(\omega) < \|W_k(\omega)\|_\infty$ for some i. If $\alpha_i(\omega)$ is varied, it will indirectly change $W_k(\omega)$ for all $k \neq i$ by changing the value of $R(\omega)$. The corresponding derivative is

$$\frac{d(\|W_k(\omega)\|_\infty^2)}{d(\alpha_i^2(\omega))} = \frac{d(\|W_k(\omega)\|_\infty^2))}{d(R^2(\omega))}\frac{d(R^2(\omega))}{d(\alpha_i^2(\omega))} = \frac{\|W_k(\omega)\|_\infty^2}{R^2(\omega)}\frac{1}{T^2}\sum_{k\neq i}\beta_k^2(\omega) > 0$$

Since the derivative is nonzero, $\|W_k(\omega)\|_\infty$ can be reduced by varying $\alpha_i(\omega)$, as long as $W_i(\omega) < \|W_k(\omega)\|_\infty$. The same holds for varying $\beta_i(\omega)$; just switch α and β in the above derivation.

We have established that

$$W_i(\omega) < \|W_k(\omega)\|_\infty \text{ for some } i \quad \Rightarrow \quad \inf_{\alpha_k(\omega),\beta_k(\omega)} W_k(\omega) \neq \|W_k(\omega)\|_\infty$$

Therefore

$$\|W_k(\omega)\|_\infty = \inf_{\alpha_k(\omega),\beta_k(\omega)} \|W_k(\omega)\|_\infty$$

$$\Rightarrow \quad W_i(\omega) = \|W_k(\omega)\|_\infty \; \forall \, i$$

$$\Rightarrow \quad W_i(\omega) = W_j(\omega) \; \forall \, i,j$$

$$\Rightarrow \quad \frac{M_i(\omega)}{\alpha_i(\omega)\beta_i(\omega)} = \frac{M_j(\omega)}{\alpha_j(\omega)\beta_j(\omega)} \; \forall \, i,j$$

$$\Rightarrow \quad \alpha_k(\omega)\beta_k(\omega) = KM_k(\omega) \; \forall \, k \text{ and for some } K > 0$$

Now re-evaluate the derivatives of R^2 when the product $\alpha_k\beta_k$ is constrained to be fixed at each k: $\alpha_k(\omega)\beta_k(\omega) = KM_k(\omega) \equiv c_k(\omega)$.

$$R^2(\omega) = \frac{1}{T^2} \sum_k \sum_{n \neq k} \alpha_k^2(\omega) \frac{c_n^2(\omega)}{\alpha_n^2(\omega)}$$

$$\begin{aligned}
\frac{d(R^2(\omega))}{d(\alpha_i^2(\omega))} &= \frac{1}{T^2} \sum_{k \neq i} \left(\frac{c_k^2(\omega)}{\alpha_k^2(\omega)} - \frac{c_i^2(\omega)\alpha_k^2(\omega)}{\alpha_i^4(\omega)} \right) \\
&= \frac{1}{T^2} \sum_{k \neq i} \left(\beta_k^2(\omega) - \frac{\beta_i^2(\omega)\alpha_k^2(\omega)}{\alpha_i^2(\omega)} \right) \\
&= \frac{1}{T^2} \frac{1}{\alpha_i^2(\omega)} \sum_{k \neq i} \left(\alpha_i^2(\omega)\beta_k^2(\omega) - \alpha_k^2(\omega)\beta_i^2(\omega) \right)
\end{aligned}$$

and by symmetry,

$$\frac{d(R^2(\omega))}{d(\beta_i^2(\omega))} = \frac{1}{T^2} \frac{1}{\beta_i^2(\omega)} \sum_{k \neq i} \left(\beta_i^2(\omega)\alpha_k^2(\omega) - \beta_k^2(\omega)\alpha_i^2(\omega) \right)$$

Since we know that for optimal α and β, $M_k^2(\omega)/\alpha_k^2(\omega)\beta_k^2(\omega)$ is constant for all k, optimality implies that

$$\frac{d(R^2(\omega))}{d(\alpha_i^2(\omega))} = \frac{d(R^2(\omega))}{d(\beta_i^2(\omega))} = 0 \quad \forall \, i$$

Define $\kappa_i(\omega) = \beta_i(\omega)/\alpha_i(\omega)$. Now pick any two integers i, j so that $i \neq j$. Optimality implies that

$$\sum_{k\neq j}\left(\alpha_j^2(\omega)\beta_k^2(\omega) - \alpha_k^2(\omega)\beta_j^2(\omega)\right) = \sum_{k\neq i}\left(\alpha_i^2(\omega)\beta_k^2(\omega) - \alpha_k^2(\omega)\beta_i^2(\omega)\right) = 0$$

Now substitute κ into the equations:

$$\alpha_j^2(\omega)\sum_{k\neq j}\beta_k^2(\omega) - \kappa_j^2(\omega)\alpha_j^2(\omega)\sum_{k\neq j}\alpha_k^2(\omega) = 0$$

$$\alpha_i^2(\omega)\sum_{k\neq i}\beta_k^2(\omega) - \kappa_i^2(\omega)\alpha_i^2(\omega)\sum_{k\neq i}\alpha_k^2(\omega) = 0$$

Divide the top equation by $\alpha_j^2(\omega)$ and the bottom by $\alpha_i^2(\omega)$. Subtract the two resulting equations to get

$$\beta_i^2(\omega) - \beta_j^2(\omega) - \kappa_j^2(\omega)\sum_{k\neq j}\alpha_k^2(\omega) + \kappa_i^2(\omega)\sum_{k\neq i}\alpha_k^2(\omega) = 0$$

$$\beta_i^2(\omega) - \beta_j^2(\omega) - \left[\kappa_j^2(\omega)\sum_{k}\alpha_k^2(\omega) - \kappa_j^2(\omega)\alpha_j^2(\omega)\right] + \left[\kappa_i^2(\omega)\sum_{k}\alpha_k^2(\omega) - \kappa_i^2(\omega)\alpha_i^2(\omega)\right] = 0$$

$$\left(\kappa_i^2(\omega) - \kappa_j^2(\omega)\right)\sum_{k}\alpha_k^2(\omega) = 0$$

$$\kappa_i^2(\omega) = \kappa_j^2(\omega)$$

Therefore, in order for $\alpha(\omega)$ and $\beta(\omega)$ to be optimal, the ratio $\beta_i(\omega)/\alpha_i(\omega)$ must be constant with respect to i. So optimality implies that

$$\alpha_i(\omega) = K_1\sqrt{M_i(\omega)}$$

$$\beta_i(\omega) = K_2\sqrt{M_i(\omega)}$$

for some nonzero real constants K_1 and K_2.

Now examine $W_i(\omega)$ when $\alpha(\omega)$ and $\beta(\omega)$ have this form.

$$W_i(\omega) = \frac{R(\omega)M_i(\omega)}{\alpha_i(\omega)\beta_i(\omega)} = \frac{R(\omega)}{K_1K_2}$$

where

$$R^2(\omega) \;=\; \frac{1}{T^2}\sum_k \sum_{n\neq k} K_1^2 M_k(\omega) K_2^2 M_n(\omega) \;=\; \frac{K_1^2 K_2^2}{T^2}\sum_k \sum_{n\neq k} M_k(\omega) M_n(\omega)$$

so

$$W_i(\omega) \;=\; \frac{1}{T}\sqrt{\sum_k \sum_{n\neq k} M_k(\omega) M_n(\omega)}$$

which is *independent of* K_1 and K_2. Therefore all solutions of this form are optimal and $\alpha(\omega)$ and $\beta(\omega)$ are optimal if and only if

$$\alpha_k(\omega) \;=\; K_1\sqrt{M_k(\omega)} \qquad \beta_k(\omega) \;=\; K_2\sqrt{M_k(\omega)}$$

for some $K_1,\ K_2 > 0$.

This concludes the proof.

The theorem shows that the optimal choice for α and β is not unique, but that all optimal solutions are related by a pair of constant scaling factors K_1 and K_2. For the sake of simplicity, we can set $K_1 = K_2 = 1$ for all $\omega \in [0, \omega_s)$ and get a simple formula for an optimal solution. This is given by the following corollary:

Corollary to Theorem 2. An optimal solution to the minimization problem

$$\inf_{\alpha,\beta} \left\| \frac{RM}{\alpha\beta} \right\|_\infty$$

is given by

$$\alpha(\omega) \;=\; \beta(\omega) \;=\; \sqrt{M(\omega)}$$

Proof: This follows directly from the theorem above when K_1 and K_2 are set to 1 for all values of $\omega \in [0, \omega_s)$.

Let's review the steps necessary to apply this result to stability analysis. Starting with a SISO hybrid feedback system of the form in Figure 3.4-1, rearrange the system components into the form of Figure 3.4-4. All the original system components are lumped into the LTI closed-loop operator

$$M \;=\; FGHD(I - \tfrac{1}{T}FGHD)^{-1}$$

which is the closed-loop transfer function when the sampler is replaced by its DC gain $1/T$. Determine the stability of this function M by some means; for example, the Nyquist test. If M is unstable, the system fails the sufficient test for stability. If M is stable, continue by computing

$$\alpha(\omega) \;=\; \beta(\omega) \;=\; \sqrt{|M(j\omega)|}$$

Now compute the radius:

$$R^2(\omega) \;=\; \frac{1}{T^2} \sum_k \sum_{n \neq k} \alpha_k^2(\omega) \beta_k^2(\omega)$$

This can be approximated numerically with relatively simple software. Chapter 4 discusses this topic in more detail. Finally, compute and plot

$$W(\omega) \;=\; \frac{R(\omega)|M(j\omega)|}{\alpha(\omega)\beta(\omega)}$$

over a sufficiently wide range of ω. If $W(\omega) < 1$ for all frequencies ω, the system passes the sufficient test for stability.

As formulated here, this process allows us to test for the stability of the *nominal* plant. This is not very exciting in itself, because there are well-known, easier ways to do this: for example, by discretizing the plant (finding the z-transform of its impulse response sampled at period T) and checking to see whether all the poles of the resulting closed-loop z-transform are inside the unit circle. The z-transform method has the further advantage of being completely nonconservative: it is a necessary and sufficient test for stability. However, by extending this conic sector test to handle the case of *robust* stability — testing stability for an entire set of plants — we can go beyond what is possible with the z-transform method. This is the topic of the next section.

3.4.2 Robust Stability for a System with One Sampler

Section 2.6 introduced the block-diagonal Δ-block model for plant uncertainty, in which an uncertain transfer function is represented by the linear fractional transformation

$$F_l(P,\Delta) \;=\; P_{11} + P_{12}(I - \Delta P_{22})^{-1}\Delta P_{21}$$

where Δ is an uncertain block-diagonal matrix with maximum gain less than one: $\|\Delta\|_\infty < 1$. In the following discussion we will define the set $\mathcal{X}_1$ by:

$$\mathcal{X}_1 \;=\; \{\Delta \mid \Delta = \mathrm{diag}(\Delta_1, \Delta_2, \ldots, \Delta_n) \;\text{ and }\; \|\Delta\|_\infty < 1\}$$

Suppose that a SISO hybrid compensator is connected in a feedback arrangement with an uncertain analog plant, and that the resulting system is modeled as in Figure 3.4-5. The single sampler closes the upper loop around P, and a diagonal uncertainty block $\Delta \in \mathcal{X}_1$ closes the lower loop. The system is said to be robustly stable if it maintains stability for all $\Delta \in \mathcal{X}_1$. Now convert Figure 3.4-5 into Figure 3.4-6 using the method of the preceding section: (a) sandwich the sampler between fictitious SISO operators α, β, and their inverses; (b) construct a conic sector $\mathrm{Cone}(1/T\beta\alpha, R)$ containing the sandwiched sampler K, where $R = \mathrm{Rad}(\alpha,\beta)$; (c) break K into an LTI operator $1/T\beta\alpha$ and a cone-bounded PLTV operator $\hat{K}$ inside $\mathrm{Cone}(0,R)$; and (d) lump the components in the LTI path together to get $1/T$.

Now we may use the conic sector stability criterion to state a two-part sufficient condition for robust stability: this system is robustly stable if (1) it is stable for all $\Delta \in \mathcal{X}_1$ when the cone center is substituted for the cone-bounded operator (equivalently, when $\hat{K}$ is set to zero) and (2) it satisfies

$$\left\| \frac{R}{\alpha\beta} F_l(P,\Delta) \left(I - \tfrac{1}{T} F_l(P,\Delta)\right)^{-1} \right\|_\infty < 1 \quad \forall\, \Delta \in \mathcal{X}_1$$

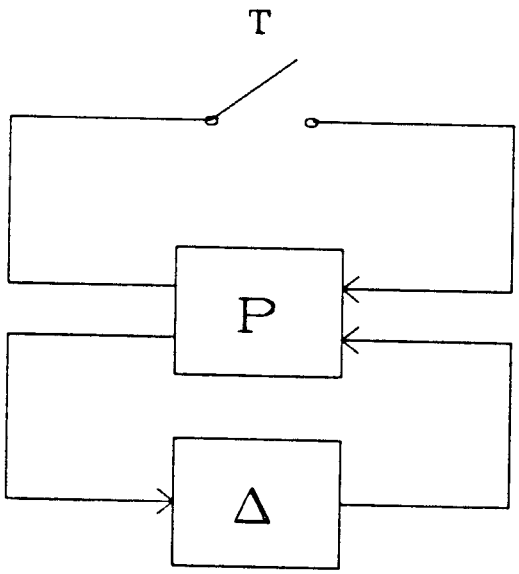

Fig. 3.4-5: A SISO sampled-data system with model uncertainty $\Delta \in \mathcal{X}_1$.

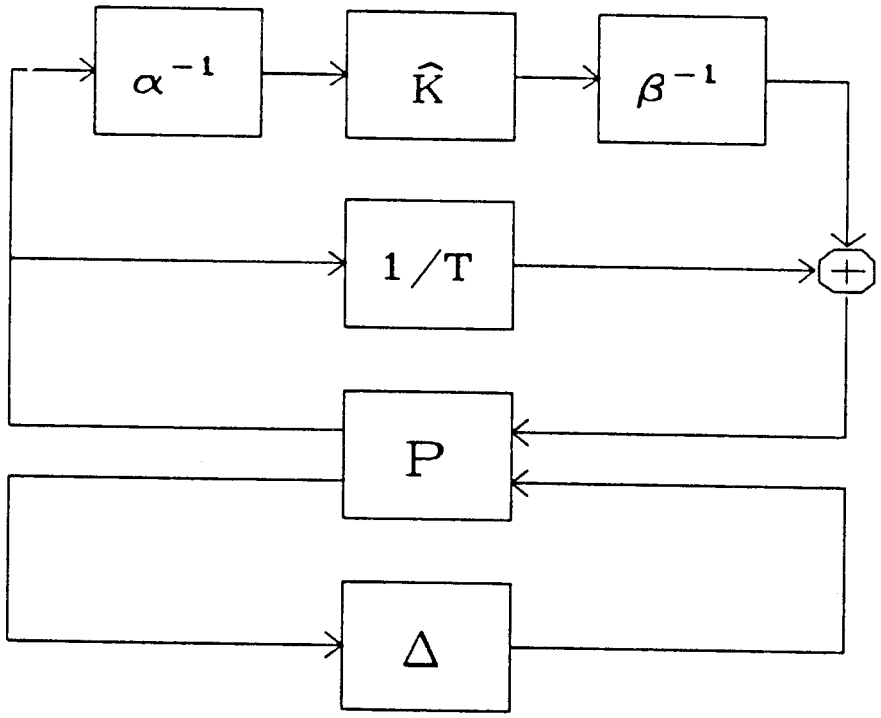

Fig. 3.4-6: The system after sandwiching the sampler with α, β and replacing the LTI path $C/\alpha\beta = 1/T$. $\hat{K}$ is inside Cone$(0, R)$.

The form of this inequality is the same as the nominal stability test of the previous section, except that the transfer function $FGHD$ is replaced by the *set* of transfer functions $F_l(P,\Delta)$, $\Delta \in \mathcal{X}_1$.

Because Δ is block-diagonal, we can rearrange Figure 3.4-6 into a simpler form. First define a PLTV operator $\hat{\Delta}$ by $\hat{K} = \hat{\Delta}R$, where $\hat{K}$ is the zero-center PLTV operator defined above, and R is the cone radius $R = \mathrm{Rad}(\alpha,\beta)$. By the elementary cone propagation results of Section 2.3.1, it is easily shown that $\hat{\Delta}$ is inside $\mathrm{Cone}(0,I)$ where I is the identity operator. Define M to be the LTI operator in the dashed box of Figure 3.4-7, containing the original plant P, the sampler's DC response $1/T$, the fictitious operators α^{-1} and β^{-1}, and the cone radius R. Now Δ and $\hat{\Delta}$ may be lumped into a single block-diagonal operator, and the entire system may be redrawn in Figure 3.4-8 as a single loop connecting M and $\mathrm{diag}(\hat{\Delta},\Delta)$. The components of M are

$$\begin{aligned}
M_{11} &= R\alpha^{-1}P_{11}\left(I-\tfrac{1}{T}P_{11}\right)^{-1}\beta^{-1} \\
M_{12} &= R\alpha^{-1}\left(I-\tfrac{1}{T}P_{11}\right)^{-1}P_{12} \\
M_{21} &= P_{21}\left(I-\tfrac{1}{T}P_{11}\right)^{-1}\beta^{-1} \\
M_{22} &= P_{22}+P_{21}\tfrac{1}{T}\left(I-\tfrac{1}{T}P_{11}\right)^{-1}P_{12}
\end{aligned}$$

Let's investigate how the robust stability conditions given above are described in this framework. First, the system must be stable for all $\Delta \in \mathcal{X}_1$ when the PLTV operator $\hat{\Delta}$ is set to zero. When this is done, the system is completely LTI and the robust stability method of Section 2.6.3 may be used. This method itself has two (necessary and sufficient) conditions: (1a) the nominal system M must be stable; and (1b) M must satisfy the inequality

$$\|\mu(M_{22})\|_\infty < 1$$

Condition (1a) may be checked by conventional means, such as the Nyquist test. An upper bound for $\mu(M_{22})$ at each frequency may be computed by the software

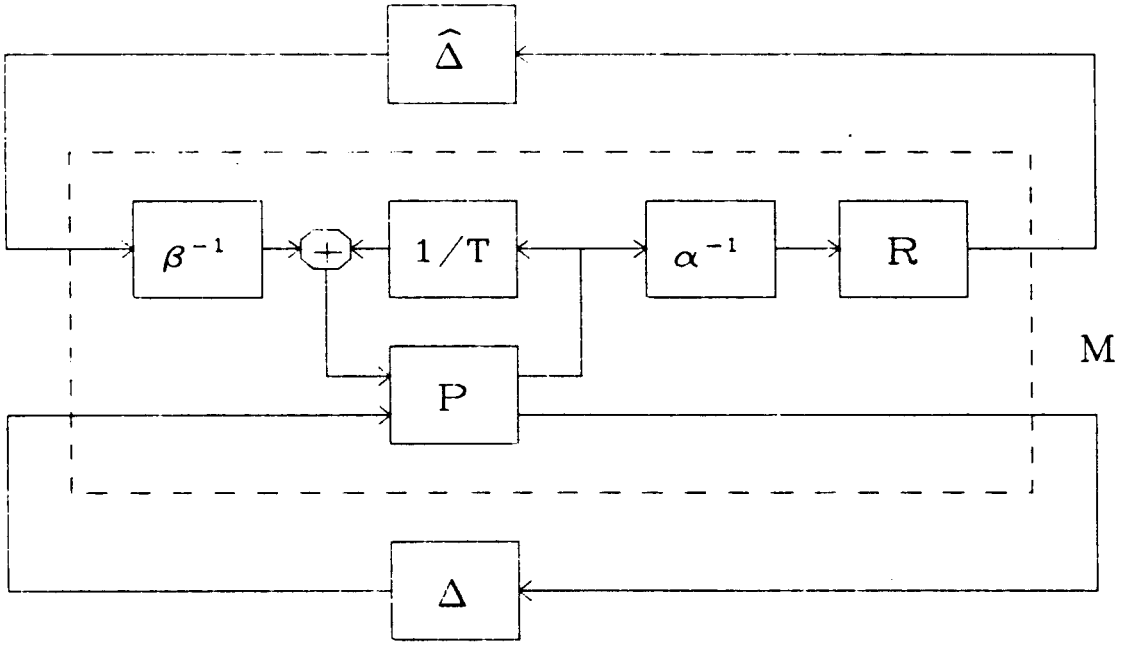

Fig. 3.4-7: $\hat{\Delta}$ is inside Cone$(0, I)$. Call the block 2×2 LTI operator in the dashed box M.

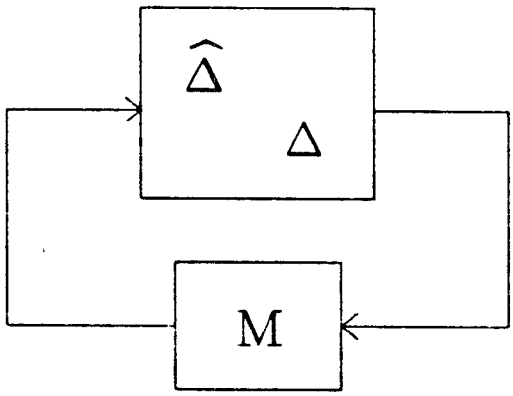

Fig. 3.4-8: Δ and $\hat{\Delta}$ can be lumped into a single block diagonal operator.

described in Chapter 4. It is easily shown using the definition for μ that $\mu(M) \geq \mu(M_{22})$; therefore $\|\mu(M)\|_\infty < 1$ is a sufficient condition for $\|\mu(M_{22})\|_\infty < 1$.

The second requirement for robust stability is

$$\left\| \frac{R}{\alpha\beta} F_l(P,\Delta) \left(I - \tfrac{1}{T} F_l(P,\Delta) \right)^{-1} \right\|_\infty < 1 \quad \forall\, \Delta \in \mathcal{X}_1$$

which, by the definition of M, is equivalent to

$$\|F_l(M,\Delta)\|_\infty < 1 \quad \forall\, \Delta \in \mathcal{X}_1$$

A sufficient condition for this is

$$\|\mu(M)\|_\infty < 1$$

Provided that M is nominally stable, this bound on $\mu(M)$ is a sufficient condition for *both* parts of the conic sector stability criterion. We have arrived at the following simple statement of a robust stability criterion:

The system in Figure 3.4-5 is stable for all $\Delta \in \mathcal{X}_1$ if

(a) M is stable and

(b) $\|\mu(M)\|_\infty < 1$

This still leaves open the question of how to make the best choice for α and β. An obvious approach to try is to find α and β to minimize $\|\mu(M)\|_\infty$. It is not so obvious how to do this, however, especially since $\mu(M)$ is not an analytic function of α and β. A more fruitful approach is to recall that the condition on $\mu(M)$ is based on simultaneously satisfying two sub-conditions: $\|\mu(M_{22})\|_\infty < 1$ and $\|F_l(M,\Delta)\|_\infty < 1$. Since M_{22} is independent of α and β, our choice for these fictitious operators will have no effect on the first of these conditions; the system will either pass this test or not, regardless of what we do. The second test is strongly reminiscent of the SISO test for nominal stability:

$$\left\| \frac{R}{\alpha\beta} FGHD \left(I - \tfrac{1}{T} FGHD \right)^{-1} \right\|_\infty < 1$$

The optimal solution in that case was to set $\alpha = \beta = \sqrt{|FGHD(I - \frac{1}{T}FGHD)^{-1}|}$. This suggests the following (suboptimal) solution to this robust stability problem: at each frequency let

$$\alpha = \beta = \sqrt{\sup_{\Delta \in \mathcal{X}_1} \left| F_l(P,\Delta)\left(I - \tfrac{1}{T}F_l(P,\Delta)\right)^{-1} \right|}$$

This supremum can be computed in the following way. Extract α^{-1}, β^{-1}, and R from the block diagram for M and add a scalar multiplier κ in the upper output channel. The resulting system is shown in Figure 3.4-9; call it $\hat{P}$. The transfer function from the upper input to the upper output, when the lower loop is closed through Δ, is given by $F_l(\hat{P},\Delta) = \kappa F_l(P,\Delta)(I - \frac{1}{T}F_l(P,\Delta)^{-1}$. If we find κ so that $\sup_{\Delta \in \mathcal{X}} F_l(\hat{P},\Delta) = 1$, then the supremum above is simply $1/\kappa$. But it can be shown, by the definition of μ, that this is equivalent to finding κ so that $\mu(\hat{P}) = 1$, since we have already assumed that $\mu(\hat{P}_{22}) = \mu(M_{22}) < 1$. (The system would fail the robust stability test if this condition were not met.) κ enters only in the top row of the matrix $\hat{P}$; therefore we can compute the supremum by the following steps:

(1) construct $\hat{P}$ with $\kappa = 1$;

(2) scale the top row of $\hat{P}$ up or down by trial values of κ until $\mu(\hat{P}) = 1$;

(3) the supremum is given by $1/\kappa$. The computation is repeated at each frequency.

This process might appear to be very computation-intensive, but programs which perform it at satisfactory speeds on a minicomputer (the VAX 11/750) have been written. A simple application of Newton's method to find the root of $1 - \mu(\hat{P})$ as κ varies gives adequate results. $\mu(\hat{P})$ is recomputed for each trial value of κ until $|1 - \mu(\hat{P})| < \epsilon$ for some small convergence test ϵ.

An alternative method for choosing α and β which is more straightforward but also more conservative is simply to pick α and β to optimize the *nominal* stability

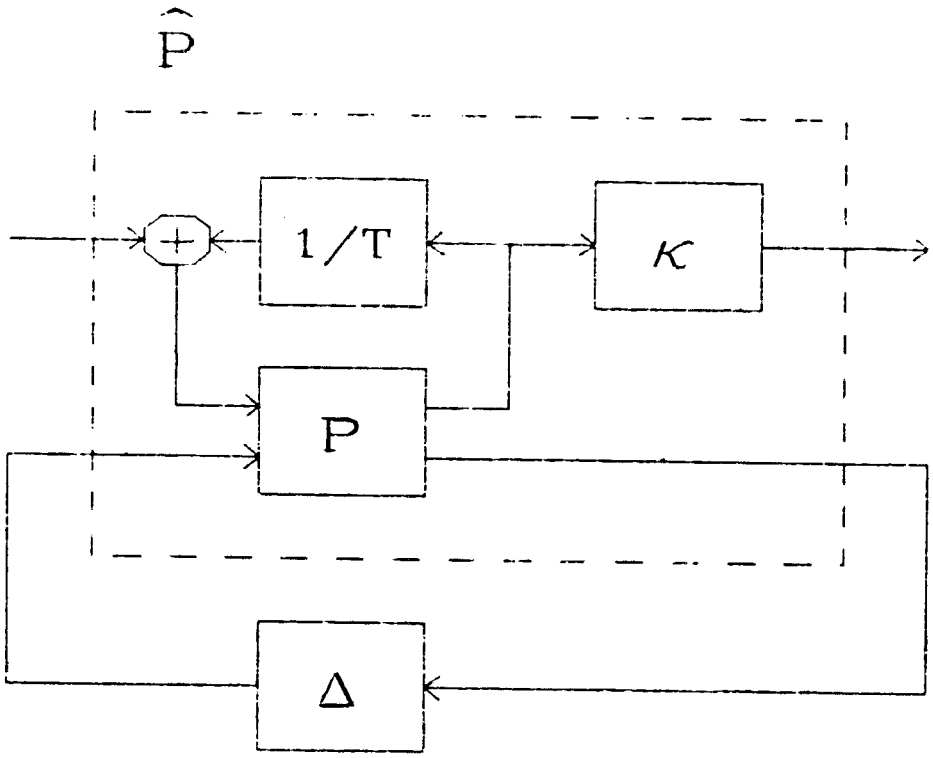

Fig. 3.4-9: The composite operator $\hat{P}$. If $\mu(P_{22}) < 1$ and $\mu(\hat{P}) = 1$, then

$$\sup_{\Delta \in \mathcal{X}_1} \left| F_l(P,\Delta)(I - \tfrac{1}{T} F_l(P,\Delta))^{-1} \right| = \frac{1}{\kappa}$$

problem; i.e., for $\Delta = 0$. Now $F_l(P, \Delta) = P_{11}$, so the solution is:

$$\alpha = \beta = \sqrt{\left| P_{11} \left(I - \tfrac{1}{T} P_{11} \right)^{-1} \right|}$$

This approach is certainly much easier than the repeated calculations of μ described above. The amount of added conservativeness it produces varies with the problem being solved. Generally speaking, if the value of $\|M_{11}\|_\infty = \| \frac{R}{\alpha\beta} P_{11} (I - \frac{1}{T} P_{11})^{-1} \|_\infty$ is not very close to 1 when these simple choices for α and β are used, then this simple method is nonconservative. This simple approach is used in the examples of Chapter 5.

3.5 Stability Criterion for a Cone-bounded Diagonal Operator

Suppose that a diagonal operator $K = \text{diag}(K_1, \ldots, K_n)$ is connected in feedback with a MIMO LTI operator G, as in Figure 3.5-1. Suppose also that each element K_i of K is known to be cone-bounded with cone center 0: K_i is inside $\text{Cone}(0, R_i)$ for $i = 1$ to n. The following lemma lets us construct a cone bound for the entire operator K out of the cone bounds for its elements.

Lemma. Let $K = \text{diag}(K_1, \ldots, K_n)$ and let each K_i be inside $\text{Cone}(0, R_i)$ as described above. Construct the LTI MIMO operator $R = \text{diag}(R_i, \ldots, R_n)$. Then K is inside $\text{Cone}(0, R)$.

Proof. Let the input to K be $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ and let its output be $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ where the elements of u and y correspond to the elements of K. Then, since K is diagonal, $y_i = K_i u_i$ for each i from 1 to n. Assume that $u \in L_{2e}$. By the definition of conic sectors,

$$\|y_i\|_\tau = \|K_i u_i\|_\tau \leq \|R_i u_i\|_\tau \quad \forall\, \tau > 0.$$

Since $\|u\|_\tau \equiv \langle u, u \rangle_\tau^{1/2}$,

$$\|y\|_\tau = \sqrt{\sum_{i=1}^{n} \|y_i\|_\tau^2} \quad \text{and} \quad \|u\|_\tau = \sqrt{\sum_{i=1}^{n} \|u_i\|_\tau^2}$$

Therefore

$$\|y\|_\tau^2 = \sum_{i=1}^{n} \|y_i\|_\tau^2 \leq \sum_{i=1}^{n} \|R_i u_i\|_\tau^2 = \|Ru\|_\tau^2 \quad \forall\, \tau > 0$$

$$\Rightarrow \quad \|y\|_\tau \leq \|Ru\|_\tau \quad \forall\, \tau > 0$$

$$\Rightarrow \quad K \text{ is inside } \text{Cone}(0, R).$$

Now a straightforward application of the conic sector stability criterion, using this new cone radius R, yields this result:

The feedback system in Figure 3.5-1 is stable if $\|RG\|_\infty < 1$.

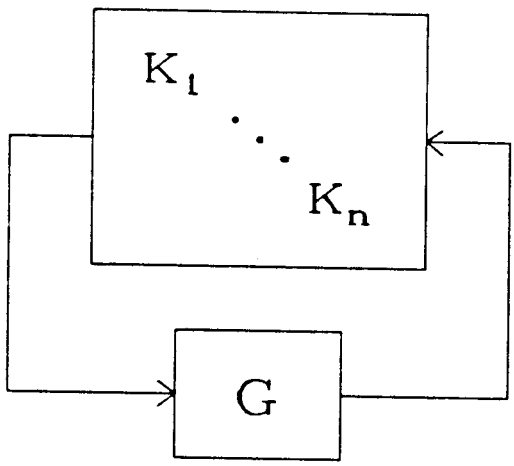

Fig. 3.5-1: Feedback system with cone-bounded diagonal operator $K = \text{diag}(K_1, \ldots, K_n)$. Each element K_i is inside $\text{Cone}(0, R_i)$.

This is a first step toward a less conservative use of conic sectors for MIMO systems. The diagonal structure of K allows us to construct a diagonal radius R which weights each input channel of K in proportion to the maximum gain of K in that channel. K must be diagonal for this to work, since the result depends on the fact that each output y_i dependent on only one input u_i.

This result is more general than it appears. Given any collection of cone-bounded operators $\hat{K}_i$ interconnected with LTI operators in an arbitrary feedback arrangement, we can always rearrange the system's block diagram so that the operators $\hat{K}_i$ form a single block-diagonal operator. This rearrangement technique was demonstrated in Section 2.5. If each $\hat{K}_i$ is inside Cone(C_i, R_i), the LTI cone centers may be extracted, resulting in new zero-center operators $K_i = \hat{K}_i - C_i$. The cone centers C_i may then be lumped in with the rest of the LTI plant, leaving the system with the structure assumed above: an operator K =diag$(K_1, \ldots, K_n)$ where each K_i is inside Cone$(0, R_i)$.

The next section uses this result to establish a conic sector analysis technique which greatly reduces conservativeness for a class of diagonal operators.

3.6 Improved Stability Criterion for a Class of Diagonal Cone-bounded Operators

Consider again the feedback system in Figure 3.5-1 of the preceding section. Suppose that each element K_i of the diagonal operator K is not only cone-bounded by Cone$(0, R_i)$, but also satisfies the following requirement: there exists an LTI operator D_i such that D_i and K_i commute, i.e., $K_i D_i = D_i K_i$. If D_i is invertible, this also implies that $D_i K_i D_i^{-1} = K_i$.

Construct the MIMO LTI operator D = diag$(D_1, \ldots, D_n)$. Since $DD^{-1} = I$, we can insert DD^{-1} into the feedback loop on each side of K without changing the

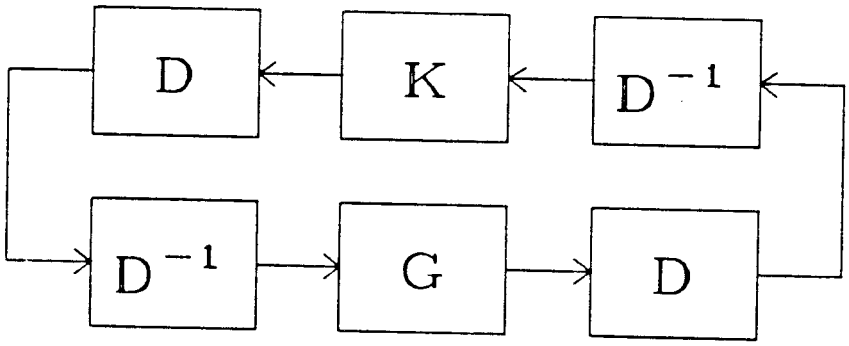

Fig. 3.6-1: System with $D^{-1}D$ pairs inserted in both sides of loop.

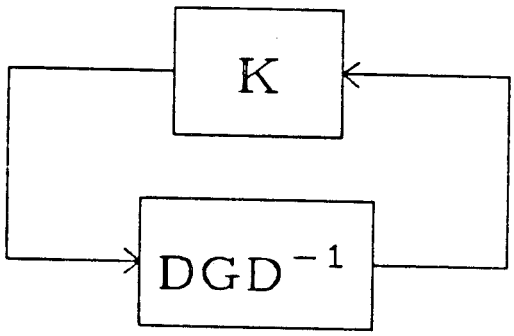

Fig. 3.6-2: Same system simplified when $K = DKD^{-1}$. With freedom to choose D, this can greatly reduce conservativeness.

operation of the system. The resulting system is shown in Figure 3.6-1. Since, by assumption, $D_i K_i D_i^{-1} = K_i$ for $i = 1$ to n, and since K and D are both diagonal, $DKD^{-1} = K$. Figure 3.6-2 results when this identity is applied to Figure 3.6-1.

By the results of the preceding section, we know that K is inside Cone$(0, R)$ where $R = \text{diag}(R_1, \ldots, R_n)$. Now the conic sector stability criterion can be applied to this modified system:

The system in Figure 3.5-1 is stable if $\|RDGD^{-1}\|_\infty < 1$.

The only requirements on D are that it be LTI and satisfy $DKD^{-1} = K$. It is not even necessary that it be diagonal. If there exists an entire set $\mathcal{D}$ of operators that meet these requirements, we can set up a minimization problem to greatly reduce the conservativeness of the conic sector method for this class of systems:

The system is Figure 3.5-1 is stable if

$$\inf_{D \in \mathcal{D}} \|RDGD^{-1}\|_\infty \; < \; 1$$

If D is restricted to be diagonal as well, then it commutes with R, which is also LTI and diagonal, so $RDGD^{-1} = DRGD^{-1}$. This form, with D and D^{-1} on the outside of the product of matrices, is reminiscent of the upper bound for μ described in Section 2.5:

$$\mu(RG) \; \leq \; \inf_{D \in \mathcal{D}} \bar{\sigma}(DRGD^{-1})$$

where in this case,

$$\mathcal{D} \; \equiv \; \{D \mid D = \text{diag}(d_1 I, \ldots, d_n I)\}$$

and where each identity matrix I has the same dimension as its corresponding Δ-block. This means that, if each operator K_i will commute with any LTI operator D_i which is a scalar times the identity, this conic sector stability test could be performed with existing μ software.

Unfortunately, the cone-bounded operators that we will use in MIMO sampled-data system analysis do not commute with *all* LTI operators of this form, but only

with a subset of them. We can sandwich each sampler in a multi-sampler system between fictitious SISO LTI operators α_i and β_i, and produce a cone bound for the resulting operator K_i using Theorem 1. But since each K_i is PLTV with period T_i, where T_i is the period of the corresponding sampler, $K_i D_i = D_i K_i$ only when D_i is a *periodic* LTI operator satisfying $D_i(j\omega) = D_i(j\omega + jn2\pi/T_i)$ for all integers n. This property of PLTV operators was shown in Section 2.7.4. It is not obvious how to solve the minimization problem when D is restricted to have periodic elements of this type.

Furthermore, this minimization problem requires that the cone radius R_i for each sandwiched sampler already be known, and this implies that the fictitious operators α_i and β_i have already been chosen. The problem of finding a good way to choose these is addressed in the next section.

3.7 A Nonconservative Conic Sector Method for MIMO, Multirate, Nonsynchronous Systems

3.7.1 New Method and Explanation

Any linear sampled-data feedback system containing multiple samplers can be represented as in Figure 3.7-1. The samplers are extracted from the system and gathered into a diagonal structure which is connected in a feedback arrangement with an $n \times n$ LTI MIMO operator M. Each element M_{ij} is the transfer function from the output of sampler j to the input of sampler i.

Now suppose we have two diagonal MIMO operators $A = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and $B = \text{diag}(\beta_1, \ldots, \beta_n)$ such that each α_i and β_i is stable and has a stable inverse. Inserting the cascades $AA^{-1} = I$ and $BB^{-1} = I$ into the loop has no effect, since any effects of A and B are cancelled by the inverses. Figure 3.7-2 shows the modified loop. Now draw a box around the cascaded connection of A, the samplers, and B, and call the resulting LTV operator K. Observe that K is diagonal. Call the period of the ith sampler T_i, and use the basic conic sector theorem for hybrid operators (Theorem 1 of Section 3.1) to compute a conic sector $\text{Cone}(C_i, R_i)$ containing each diagonal element K_i of K:

$$
\begin{aligned}
C_i(j\omega) &= \frac{1}{T_i}\beta_i(j\omega)\alpha_i(j\omega) \\
|R_i(j\omega)|^2 &= |\text{Rad}(\alpha_i, \beta_i)|^2 \\
&= \frac{1}{T_i^2}\sum_k \sum_{n \neq k} \left|\alpha_i\left(j\omega - jk\frac{2\pi}{T_i}\right)\right|^2 \left|\beta_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\right|^2
\end{aligned}
$$

As we have noted earlier, the magnitude of the radius $R_i(j\omega)$ is periodic in ω with period $2\pi/T_i$. The formula above specifies only the magnitude of R_i; a minimum phase transfer function $R_i = \text{Rad}(\alpha_i, \beta_i)$ always exists which is stable, has a stable inverse, and has the magnitude given above. In practice, we do not need to actually compute this minimum phase factorization, since only the magnitude of $R_i(j\omega)$ is

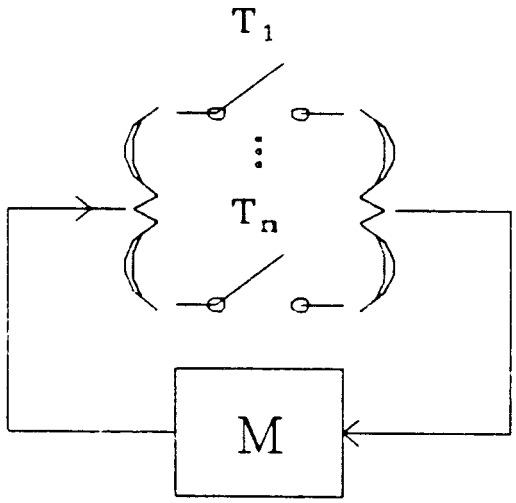

Fig. 3.7-1: A MIMO sampled-data loop. All holds, prefilters, digital filters, and other LTI components are collected in the $n \times n$ transfer function M.

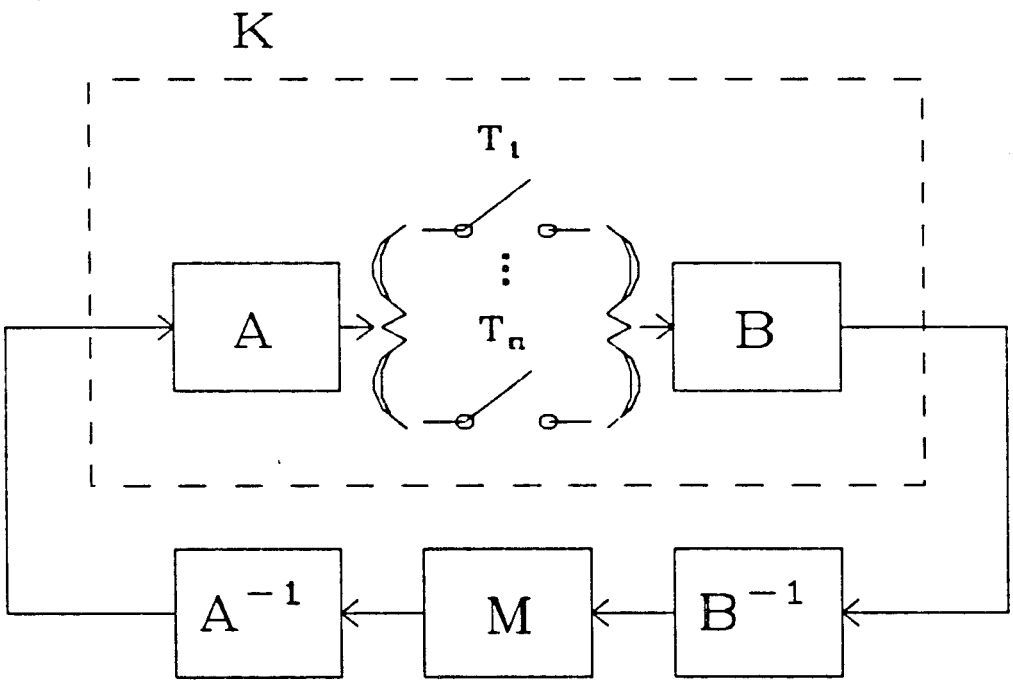

Fig. 3.7-2: The system after inserting diagonal LTI operators A, A^{-1}, B, B^{-1}, all of which are stable. The PLTV operator K can be cone-bounded with Theorem 1.

used in the stability test. If such a factorization were required, however, the complex cepstrum matching technique of Chapter 6 could be used to find a rational function in z, of arbitrary order, to approximate $|R_i(j\omega)|$.

So far, we have a separate cone for each element of K. Since K is diagonal, we can combine these into a single cone containing K by the simple technique of Section 3.5: define $C = \text{diag}(C_1, \ldots, C_n)$ and $R = \text{diag}(R_1, \ldots, R_n)$. Then K is strictly inside $\text{Cone}(C, R)$.

At this point, we can apply the conic sector stability criterion to get a sufficient condition for closed-loop stability. Define $\hat{T} = \text{diag}(T_1, \ldots, T_n)$, so that $C = \hat{T}^{-1}BA$. The system is stable if it is stable when K is replaced by C and if

$$\begin{aligned} &\|RA^{-1}MB^{-1}(I - \hat{T}^{-1}BAA^{-1}MB^{-1})^{-1}\|_\infty \\ &= \|RA^{-1}M(I - \hat{T}^{-1}M)^{-1}B^{-1}\|_\infty \\ &= \|RA^{-1}\tilde{M}B^{-1}\|_\infty \;<\; 1 \end{aligned}$$

where $\tilde{M}$ is defined by

$$\tilde{M} \;\equiv\; M(I - \hat{T}^{-1}M)^{-1}$$

The first part of the test, that the system must be stable when K is replaced by C, may be performed by standard LTI techniques. Since A and B are cancelled by their inverses, this condition is equivalent to requiring that the LTI MIMO system with loop gain $\hat{T}^{-1}M$ be closed-loop stable. This may be verified by the MIMO Nyquist test: plot $\det(I - \hat{T}^{-1}M(s))$ as s traverses a standard Nyquist D-contour in the right half-plane, in a clockwise direction. The system is stable if the number of counter-clockwise encirclements of the origin by $\det(I - \hat{T}^{-1}M)$ equals the number of unstable poles of M. Note that A and B do not appear in this part of the test.

Now let us introduce a small change in the stability test for reasons that will be explained shortly. Since each radius $R_i(j\omega)$ is a periodic LTI operator with period $2\pi/T_i$, it *commutes* with its respective sampler; this property of periodic

transfer functions was shown in Secton 2.7.4. We will exploit this property to gain a slight improvement in nonconservativeness. Factor each $R_i(j\omega)$ into minimum phase square roots: $R_i(j\omega) = R_i^{1/2}(j\omega)R_i^{1/2}(j\omega)$. This is always possible for the same reason that a minimum phase $R_i(j\omega)$ can always be constructed. Again, we will not actually need to know the phase of $R^{1/2}(j\omega)$: only its magnitude is used. Let $R^{1/2} = \text{diag}(R_1^{1/2}, \ldots, R_n^{1/2})$.

Insert the cascaded pair $R^{1/2}R^{-1/2} = I$ into the system between A and A^{-1} as shown in Figure 3.7-3. Since $R^{1/2}$, A, and B are all diagonal and LTI, they commute with each other. Now "push" $R^{1/2}$ through A, the samplers, and B to produce the equivalent system in Figure 3.7-4. The conic sector stability criterion now implies stability if

$$\begin{aligned}
&\|R(R^{-1/2}A^{-1}MB^{-1}R^{-1/2}(I - \hat{T}^{-1}R^{-1/2}A^{-1}MB^{-1}R^{1/2})^{-1})\|_\infty \\
&= \|R^{1/2}A^{-1}M(I - \hat{T}^{-1}M)^{-1}B^{-1}R^{1/2}\|_\infty \\
&= \|R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}\|_\infty \;<\; 1
\end{aligned}$$

This condition differs from the previous one by splitting R into equal parts on both sides of the matrix product. This is desirable in order to preserve balancing in the matrix product. This balancing is achieved in the first place by a scaling process, which will now be explained.

So far, we have derived a MIMO conic sector stability test with fictitious operators A and B, but no method has been given to choose A and B. In the SISO case, the criterion above reduces to $\left\|\frac{R\tilde{M}}{\alpha\beta}\right\|_\infty$. Section 3.4.1 showed that an optimal choice for α and β is $|\alpha(j\omega)| = |\beta(j\omega)| = \sqrt{|\tilde{M}(j\omega)|}$. A similar choice for A and B in the MIMO case provides a suboptimal, but usually nonconservative, result.

At each frequency ω, find $D(\omega)$ to solve

$$\hat{\mu} \;=\; \inf_{D\in\mathcal{D}} \bar{\sigma}(D\tilde{M}D^{-1})$$

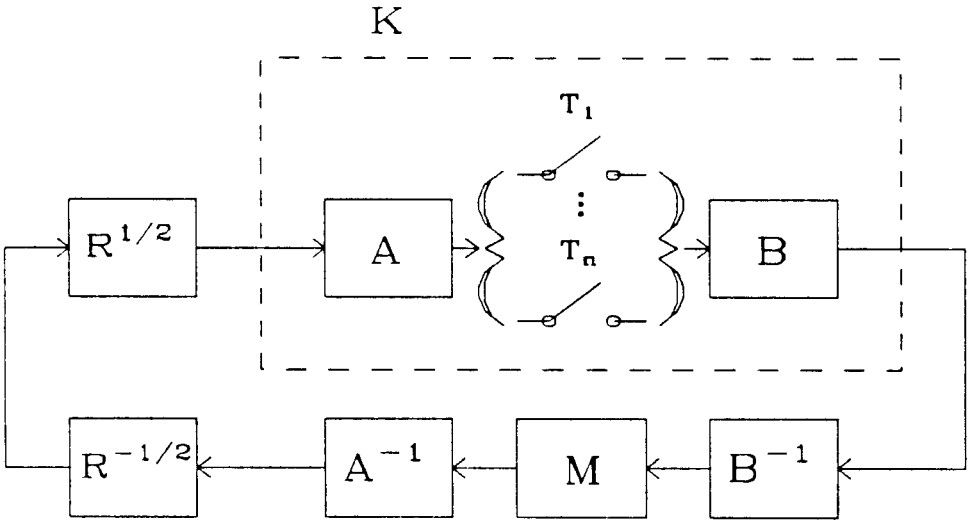

Fig. 3.7-3: Insertion of minimum phase $R^{1/2}$ and $R^{-1/2}$, where $R = \text{diag}(R_1, \ldots, R_n)$ and $R_i = \text{Rad}(\alpha_i, \beta_i)$.

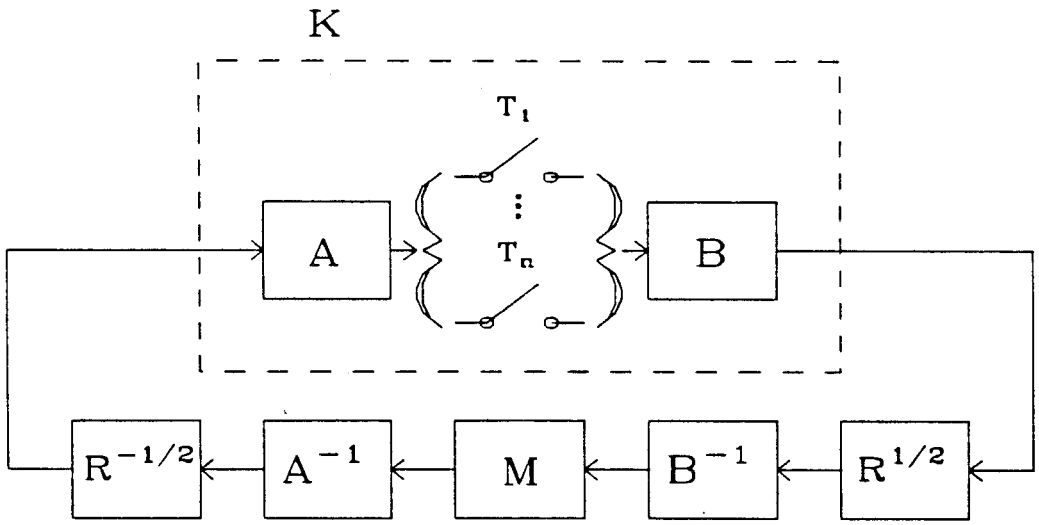

Fig. 3.7-4: $R^{1/2}$ commutes with K, since both are diagonal and each element $R^{1/2}$ is periodic in ω with period $2\pi/T_i$. This preserves balancing in the stability test $\|R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}\|_\infty < 1$.

where $\mathcal{D}$ is the set of diagonal matrices with real, positive diagonal elements. Since $D\tilde{M}D^{-1}$ is unchanged when D is multiplied by a scalar, normalize D by scaling it so that $\bar{\sigma}(D)\ \underline{\sigma}(D) = 1$. In other words, scale D so that the geometric mean of its largest and smallest diagonal elements is 1. We denote the minimum value of $\bar{\sigma}(D\tilde{M}D^{-1})$ by $\hat{\mu}$ to emphasize its close relationship to the Structured Singular Value $\mu(\tilde{M})$. Actually, $\hat{\mu}$ is an upper bound for μ [5].

Finally, let

$$A = \hat{\mu}^{1/2}D^{-1} \quad \text{and} \quad B = \hat{\mu}^{1/2}D$$

This defines each fictitious operator α_i and β_i as a positive real-valued function of frequency. To be rigorous, we should replace the real-valued $\alpha_i(j\omega)$ and $\beta_i(j\omega)$ with minimum phase functions of the same magnitude, since A and B must be stable transfer functions with stable inverses. However, since $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ is unchanged when any element of A, B, or $R^{1/2}$ is rotated in phase, we can avoid useless complexity by simply writing A and B as the real-valued functions above.

To see why these are good choices for A and B, let's see what happens when we insert them into the stability test:

$$\begin{aligned}
\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}) &= \bar{\sigma}(R^{1/2}D\hat{\mu}^{1/2}\tilde{M}D^{-1}\hat{\mu}^{-1/2}R^{1/2}) \\
&= \hat{\mu}^{-1}\bar{\sigma}(R^{1/2}D\tilde{M}D^{-1}R^{1/2}) \\
&\leq \hat{\mu}^{-1}\bar{\sigma}(R^{1/2})\bar{\sigma}(D\tilde{M}D^{-1})\bar{\sigma}(R^{1/2}) \\
&= \bar{\sigma}^2(R^{1/2}) = \bar{\sigma}(R)
\end{aligned}$$

therefore

$$\|R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}\|_\infty \leq \|R\|_\infty = \max_i \|R_i\|_\infty$$

So the stability criterion is bounded above by the largest magnitude of any of the radius functions R_i, which are all *periodic* in ω. The magnitude of each radius function $R_i = \mathrm{Rad}(\alpha_i, \beta_i)$ is determined primarily by the high-frequency roll-off

properties of α_i and β_i. The faster α_i and β_i roll off, the smaller R_i becomes. If we chose α_i and β_i to roll off faster than $\hat{\mu}^{1/2}D^{-1}$ and $\hat{\mu}^{1/2}D$, then the radii R_i would become smaller, but $\bar{\sigma}(A^{-1}\tilde{M}B^{-1})$ would grow without bound with increasing ω. If we chose α_i and β_i to roll off slower than $\hat{\mu}^{1/2}D^{-1}$ and $\hat{\mu}^{1/2}D$, then $\bar{\sigma}(A^{-1}\tilde{M}B^{-1})$ would shrink to zero with increasing ω, but the radius R_i would become larger, causing $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ to increase at low frequencies. The choice for A and B presented here strikes a balance between these two extremes, resulting in a value for $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ which neither shrinks nor grows with increasing ω.

Let's now review the steps involved in this method:

1. Draw the system as in Figure 3.7-1, with all samplers extracted into a diagonal structure with interconnection matrix M.

2. Compute $\tilde{M} = M(I - \hat{T}^{-1}M)^{-1}$, where $\hat{T} = \text{diag}(T_1, \ldots, T_n)$ and each T_i is the period of the ith sampler.

3. Perform the MIMO Nyquist test by plotting $\det(I - \hat{T}^{-1}M)$ over a Nyquist D-contour and counting the counter-clockwise encirclements of the origin. If the number of encirclements equals the number of unstable poles of M, proceed. If not, the system fails the *sufficient* condition for stability.

4. At each frequency ω, compute $\hat{\mu} = \inf_{D \in \mathcal{D}} \bar{\sigma}(D\tilde{M}D^{-1})$, where $\mathcal{D}$ is the set of positive definite real diagonal matrices. The issue of how to write software to do this is addressed in Chapter 4.

5. Scale the diagonal matrix D which achieves $\hat{\mu}$ above so that $\bar{\sigma}(D)\underline{\sigma}(D) = 1$ for all frequencies. This minimizes the radii R_i in step 7, subject to the constraint that $\hat{\mu} = \bar{\sigma}(D\tilde{M}D^{-1})$.

6. Let $A = \hat{\mu}^{1/2}D^{-1}$ and $B = \hat{\mu}^{1/2}D$, where $A = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and $B = \text{diag}(\beta_1, \ldots, \beta_n)$.

7. Compute $R^{1/2} = \text{diag}(R_1^{1/2}, \ldots, R_n^{1/2})$, where $R_i = \text{Rad}(\alpha_i, \beta_i)$.

8. Compute and plot $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$. If this function is less than 1 at

all frequencies, the system is stable.

This technique, while still suboptimal, is relatively simple and gives nonconservative results for a wide class of sampled-data systems. Since there is no restriction on the sampling periods T_i, it works equally well for multirate and single-rate systems, even when the sampling periods are not related by rational numbers. Thus the technique can be used for hybrid systems containing two computers which are not run from a common clock. There are very few known stability tests which are valid for multirate systems with sampling periods related by irrational numbers.

The new technique also eliminates nearly all the other limitations of the original technique based on Theorem 1 of Section 3.1. The compensator need not be stable, since it is now outside the cone-bounded operator. The use of diagonal scaling to balance the matrix $\tilde{M}$ can provide a vast improvement for systems with high condition number (MIMO systems with much more gain for inputs in some vector "directions" than in others). Since only the SISO version of Theorem 1 is used, the inherent conservativeness of the formula for the radius R_i is avoided. This conservativeness occurs because the MIMO version of the formula uses products of maximum singular values of matrices, thereby losing directionality information.

3.7.2 How Much Conservativeness is Left?

Both the SISO and MIMO analysis techniques proposed in this chapter take advantage of the same idea: embedding each sampler between fictitious LTI operators α_i and β_i. By choosing α_i and β_i to cause the final stability criterion $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ to be "flat" at high frequencies, that is, neither to grow nor to decay as ω increases, we eliminate much of the conservativeness of the conic sector approach. How much conservativeness is left? Are there systems for which these methods still will not work, and if so, how can we identify them?

Four situations are identified below in which the conic sector approach still gives conservative results, even when the improved methods outlined so far in this thesis are used. These are:

1. Systems unstable with cone center, that is, unstable when each sampler is replaced by $1/T_i$, where T_i is its period.

2. Cases where the optimal radius is too large.

3. Systems with significant skewed sampling effects, where it makes a difference whether all samplers "fire" simultaneously or with a fixed time offset.

4. Systems with a sampler at the output of a ZOH.

The first three situations generally occur when the system has a substantial amount of aliasing, caused by inadequate high-frequency rolloff. Therefore, it can be argued that most systems which fall into the first three classes are poorly designed, since their poor anti-aliasing filtering will generally cause them to have high sensitivity to noise. Furthermore, two of these situations occur only when a system is close to the border between stability and instability, indicating poor robustness to model uncertainty.

The fourth conservative situation occurs when a system has a sampler at the output of a ZOH, and the ratio of the sampler's period and the ZOH period is

not an integer. This can occur in practice when a signal is sampled at one rate, held in a buffer register, and then read at a different rate by the control software. The problem arises when both samplers are extracted from the system and put in separate conic sectors. Since the transfer function from the first sampler to the second is a ZOH, it rolls off approximately as $1/\omega$ as $\omega \to \infty$. This slow rate of rolloff causes the cone radius to be large, leading to conservativeness. The solution to this problem is to put *both* samplers and the ZOH together in a *single* conic sector, using a modified version of Theorem 1. This new theorem is presented in Section 3.8.

3.7.2.1 Systems Unstable With Cone Center

Some systems which are in fact stable will become unstable when each sampler is replaced by its DC gain $1/T_i$, where T_i is the sampler's period. This is equivalent to having $\tilde{M} = M(I - \hat{T}^{-1}M)^{-1}$ unstable. When this occurs, the conic sector stability analysis is stopped before it begins, since the system fails at least one of the two sufficient conditions for stability. The following example, due to Doyle [4], illustrates this.

Consider the hybrid feedback system in Figure 3.7-5. Let the sampling period be $T = 2\pi$ seconds, and let

$$G(s) = \frac{.45}{(s + .01)(s + .5)(s + 2)}$$

The discretized system is stable, but when the sampler is replaced by $1/T$, it is not. Figure 3.7-6 illustrates this by showing partial log Nyquist plots of $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$. The discretized loop gain $(GH_0^T)^T$ is plotted only from DC to half the sampling frequency, since it is periodic beyond that point. Both plots come very close to the critical point $s = -1$, but on opposite sides; therefore the discrete system has no encirclement of $s = -1$, and the system with cone center has one

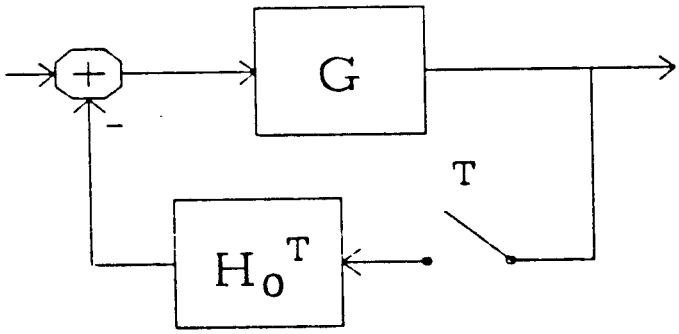

Fig. 3.7-5: This system is stable, but becomes unstable when the sampler is replaced by $1/T$. $G(s) = .45/((s+.01)(s+.5)(s+2))$ and $T = 2\pi$ seconds.

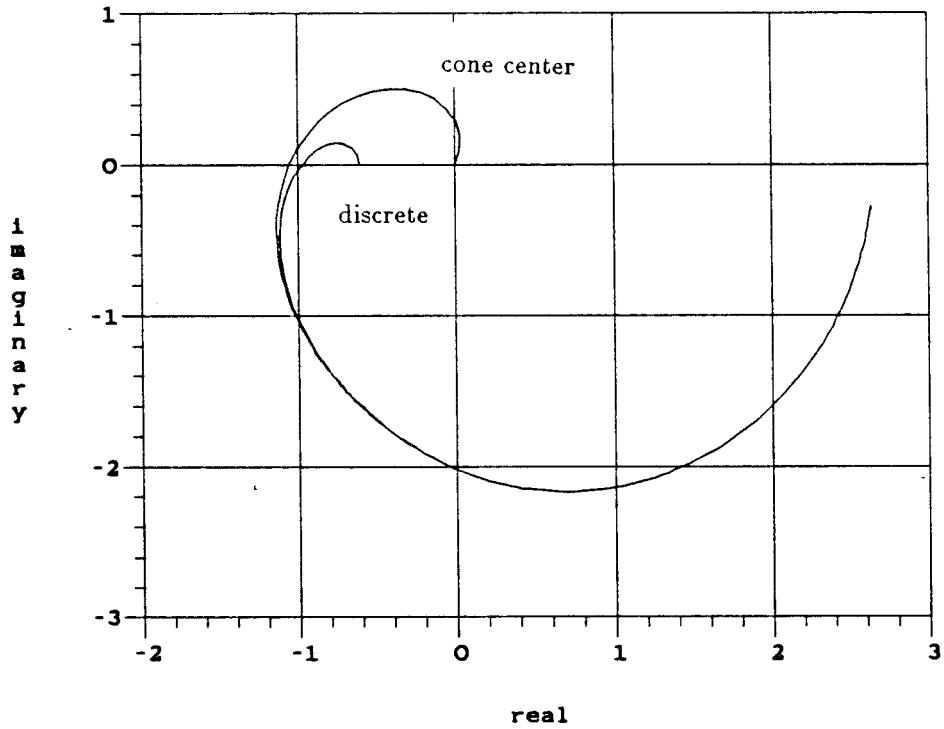

Fig. 3.7-6: Log Nyquist plots of $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$. The discrete-time loop does not encircle -1, but the LTI loop does.

encirclement, indicating instability.

The log Nyquist plot is very useful and will be used several times in this thesis. If the function being plotted is $f(j\omega) = |f(j\omega)|e^{j\phi(j\omega)}$, then the log Nyquist plot is the locus over ω of the function $(1 + \log_{10}|f(j\omega)|)e^{j\phi(j\omega)}$ for values of ω where $|f(j\omega)| \geq 0.1$. The distance from the origin is on a logarithmic scale, with 1 corresponding to $|f(j\omega)| = 1$, 2 corresponding to $|f(j\omega)| = 10$, etc. All values of $|f(j\omega)|$ less than 0.1 in magnitude are mapped into the origin. The critical point is unchanged; it is still -1. Unlike the standard linear Nyquist plot, the log Nyquist plot clearly shows a function's behavior at all frequencies, even if it has a dynamic range of several decades.

Let's try to understand Figure 3.7-6. It is well known that the discrete transfer function is related to the cone center transfer function as follows:

$$\begin{aligned}(GH_0^T)^T(e^{j\omega T}) &= \frac{1}{T}\sum_k G(j\omega - jk\omega_s)H_0^T(j\omega - jk\omega_s) \\ &= \frac{1}{T}GH_0^T + \frac{1}{T}\sum_{k\neq 0} G(j\omega - jk\omega_s)H_0^T(j\omega - jk\omega_s)\end{aligned}$$

where $\omega_s = 2\pi/T$. Therefore the discretized loop gain is equal to the cone center loop gain $\frac{1}{T}GH_0^T$ plus the sum of all its out-of-band aliased values. The smaller the out-of-band terms are, the closer the cone center and discretized loop gains are to each other. This means that a system with steep anti-aliasing filtering will show good agreement between its cone center and discretized loop gains.

However, this system *does* have steep anti-aliasing filtering: GH_0^T rolls off as $1/\omega^4$ as $\omega \to \infty$. But $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$ do show close agreement over most of the log Nyquist plot; it is a relatively small difference between them near the critical point that causes the problem. This small difference is enough to push the system into instability because the system is not robust in the first place: the discrete system has a phase margin of only 2.4°.

We may conjecture, then, that this problem will occur only for systems which

either:

a) have poor anti-aliasing filtering, and thereby have high noise sensitivity; or

b) have poor robustness margins (small gain or phase margins).

3.7.2.2 Optimal Radius Too Large

There is another class of systems which are stable in fact and stable when their samplers are replaced by $1/T_i$, but which fail the second part of the stability test even when the optimal radius is used. The following example, also due to Doyle [4], illustrates this.

Let the system be described by Figure 3.7-5 as in the previous example, but let

$$G(s) = \frac{100\pi s + 100}{100\pi k s^2 + (\pi + 100)ks + k - 100}$$

where $T = 2\pi$ seconds, as before, and $k = 2.4$. This system has one open-loop unstable pole. The closed loop system is stable, both in fact (discretized analysis) and when the sampler is replaced by $1/T$. This can be seen in Figure 3.7-7, which shows linear Nyquist plots of both $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$. The discretized loop gain is plotted only from DC to π/T, as in the previous example. Both Nyquist plots show one counter-clockwise encirclement of the critical point $s = -1$, indicating stability.

However, when the optimal radius is plotted in Figure 3.7-8, it has maximum amplitude greater than 1: about 1.15. Therefore this system fails the second sufficient condition for stability, even though it is stable. Let's examine the reasons for this.

$G(s)$ rolls off as $1/\omega$ at high frequencies, so $\tilde{M} = GH_0^T(I - \frac{1}{T}GH_0^T)^{-1}$ rolls off as $1/\omega^2$. This rate of rolloff is somewhat slow, but not atypical of some real systems. The major reason for the method's failure here is the closeness of $\frac{1}{T}GH_0^T$ to the critical point at *low* frequencies: at DC, $\frac{1}{T}GH_0^T \approx -1.03$. In a typical feedback

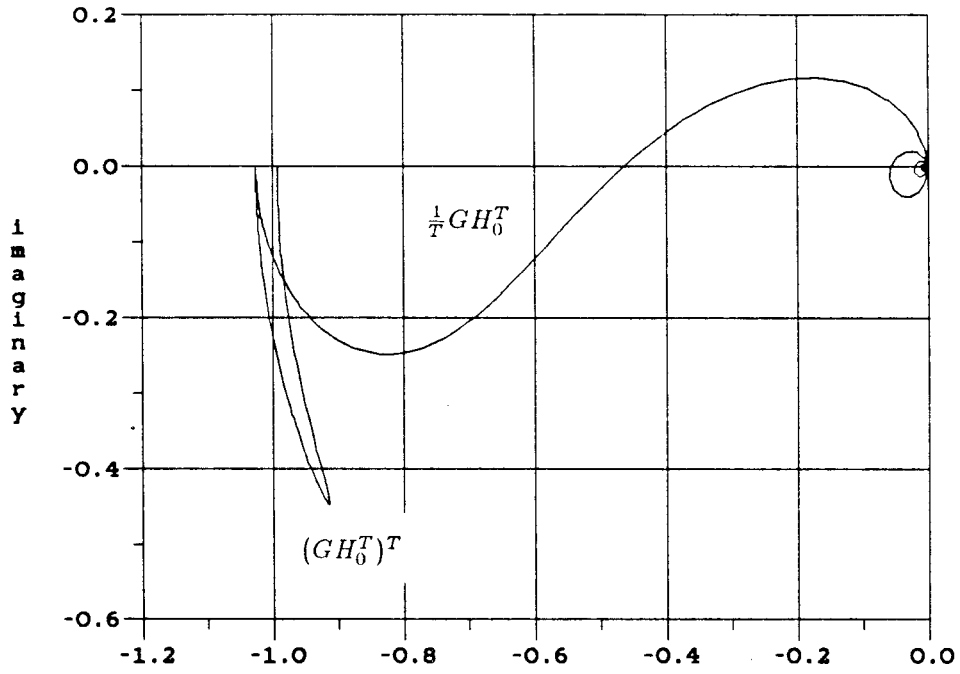

Fig. 3.7-7: With $G(s) = (100\pi s + 100)/(100\pi k s^2 + (\pi + 100)ks + k - 100)$, $k = 2.4$, $T = 2\pi$ seconds, the system in Fig. 3.7-5 is stable, but its optimal radius is greater than 1. These are linear Nyquist plots of $\frac{1}{T}GH_0^T$ and $(GH_0^T)^T$.

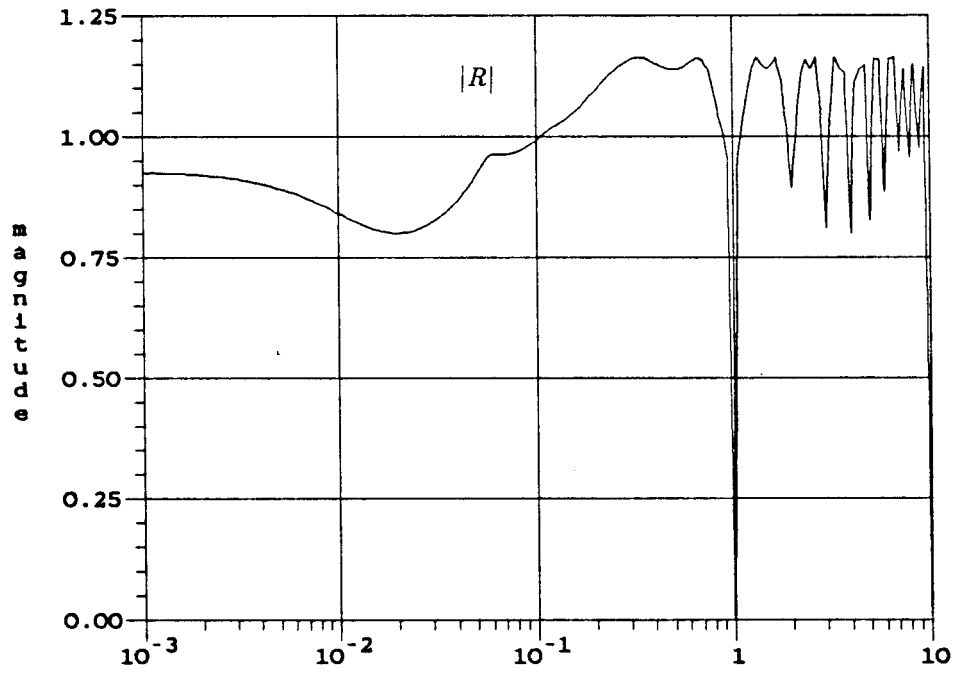

Fig. 3.7-8: Optimal radius has maximum gain $\|R\|_\infty = 1.15 > 1$.

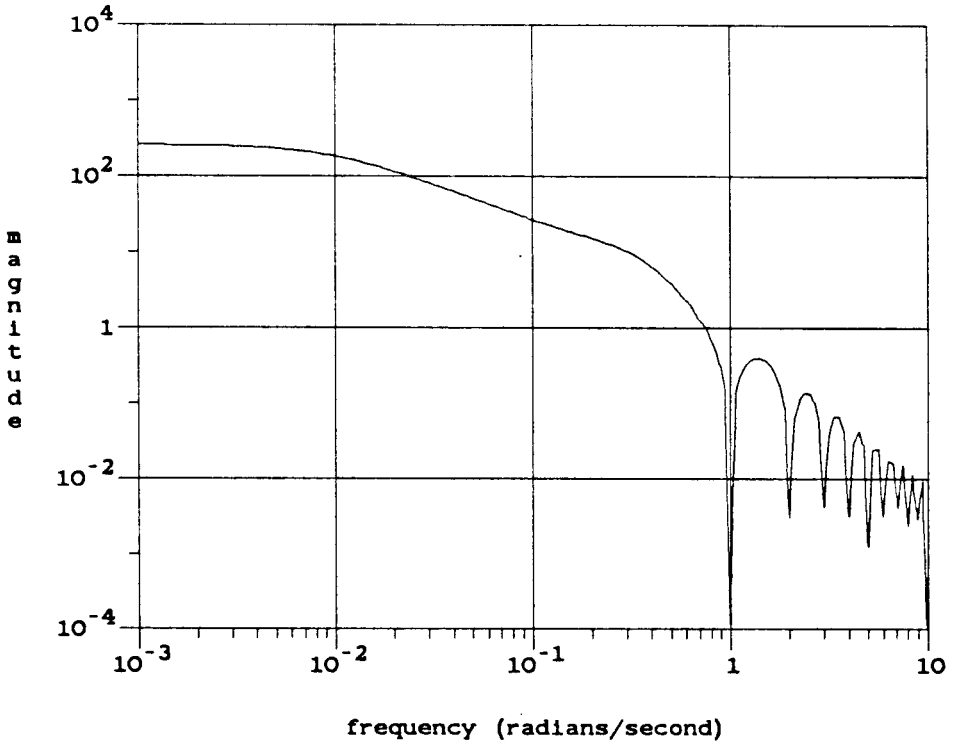

Fig. 3.7-9: Bode plot of $\tilde{M} = GH_0^T(1 - \frac{1}{T}GH_0^T)^{-1}$.

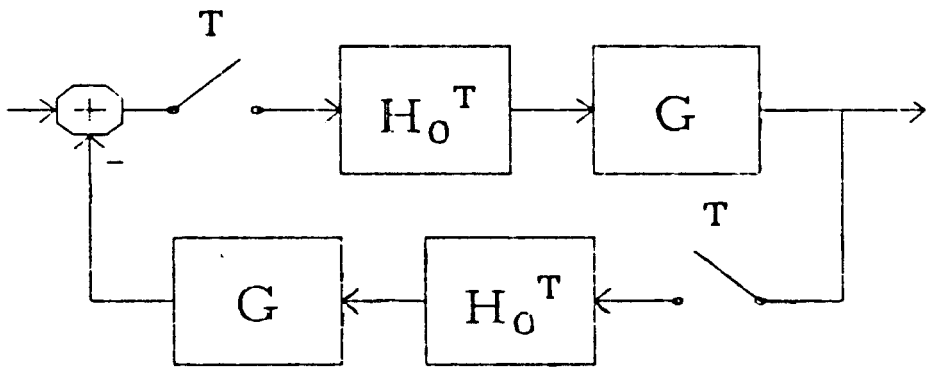

Fig. 3.7-10: This two-sampler system is unstable when the samplers fire simultaneously, but stable when sampling times are skewed by $T/2$. $G(s) = .2/(s + .1)$.

system, $|\frac{1}{T}GH_0^T| \gg 1$ at low frequencies, so $\tilde{M} \approx T$ in that frequency range. In this example, however, $\tilde{M} \approx 220 \approx 35T$ at low frequencies: see the plot of $\tilde{M}$ in Figure 3.7-9. Since $|\alpha| = |\beta| = |\tilde{M}|^{1/2}$, this makes α and β large as well, and produces a large radius, even in the optimal case.

The closeness of $\frac{1}{T}GH_0^T$ to the critical point is an indication of a poor robustness margin. In fact, the system's gain reduction margin is only 0.25 dB: if the system's gain is reduced by only that very small amount, it becomes unstable. In this example, then, as in the previous one, the system's failure of the conic sector stability test is caused by its poor robustness margin.

3.7.2.3 Skewed Sampling Effects

A system with two or more samplers running at the same rate can cause conic sector analysis to exhibit another type of conservativeness, unseen in the single-sampler case. This conservativeness is due to the fact that, when each "sandwiched" sampler is placed in a cone, its exact sampling *rate* is taken into account, but the exact *timing* of the sampler is ignored. For example, if two samplers have the same period T, the first sampler might "fire" at times $t = 0, T, 2T, \ldots$ while the second fires at $t = 0.5T, 1.5T, 2.5T, \ldots$. The amount of fixed offset, or skew, in their relative timing could have some effect on system stability, but it is ignored in the conic sector treatment.

The following example illustrates this. Figure 3.7-10 shows a system with two samplers with the same period $T = 2\pi$ seconds. The transfer function $G(s) = .2/(s + .1)$ appears twice in the system, and is preceded in both cases by a ZOH.

Since there is no model uncertainty in the system, we can perform an exact stability analysis by standard z-transform techniques. For the case when both samplers fire simultaneously, simply discretize the two (identical) paths between

the samplers to get the discrete-time loop gain:

$$L_1(z) = ((GH_0^T)^T)^2$$

Now suppose the timing of the samplers is skewed by $T/2$, half a sampling period. This may be modelled by sandwiching one of the samplers between a time delay $e^{-sT/2}$ on one side and a time advance $e^{sT/2}$ on the other, and then assuming that both samplers fire simultaneously. This produces the discrete-time loop gain for the skewed case:

$$L_2(z) = (GH_0^T e^{-sT/2})^T (GH_0^T e^{sT/2})^T$$

Finally, we may produce the loop gain with cone centers by replacing both samplers with $1/T$:

$$L_3(s) = \frac{1}{T^2}(G(s)H_0^T(s))^2$$

Figure 3.7-11 is a linear Nyquist plot of all three loop gains. The two discrete-time functions are plotted only from DC to π/T, as usual. It shows that both the cone center and the skewed-sampling loop gains do not encircle the critical point $s = -1$, so the system is stable in those two cases. However, the discrete-time loop gain for synchronized sampling does make one encirclement, so the system is unstable when the samplers are synchronized.

We are able to use the SISO Nyquist test on the cone center response, even though there are two samplers, because both samplers appear in a single SISO loop. In general, though, applying the conic sector stability test to a multi-sampler system requires use of the MIMO Nyquist test. Both samplers are extracted from the system and placed in a diagonal structure in a feedback arrangement with the 2×2 MIMO LTI operator M. For this system,

$$M = \begin{bmatrix} 0 & -GH_0^T \\ GH_0^T & 0 \end{bmatrix}$$

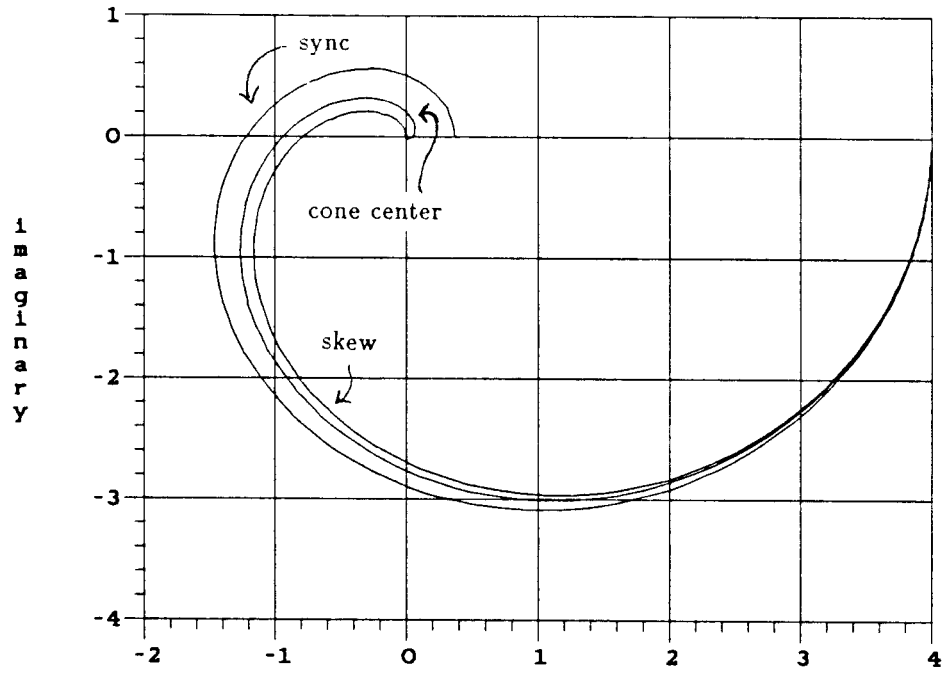

Fig. 3.7-11: Linear Nyquist plot of loop gain for a discrete-time system with and without skew, and for a system with samplers replaced by $1/T$. In the synchronized case, the loop gain encircles -1.

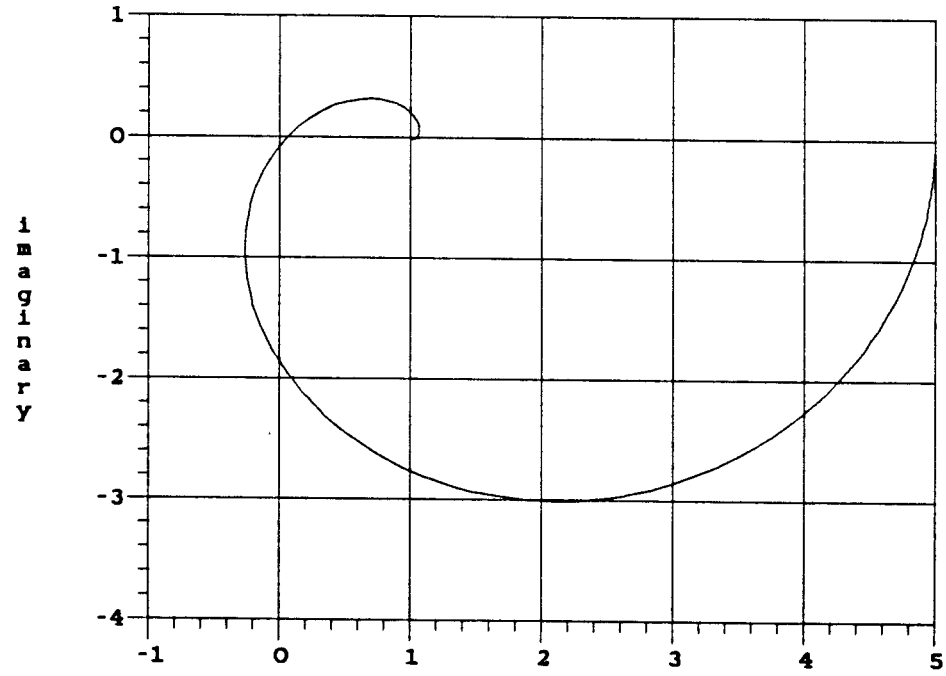

Fig. 3.7-12: MIMO Nyquist test for the same system. This is a linear Nyquist plot of $\det(I - \hat{T}^{-1}M)$. Locus does not encircle origin, so the system is stable with cone center.

To perform the MIMO Nyquist test, compute $\det(I - \hat{T}^{-1}M)$ where $\hat{T} = \mathrm{diag}(T, T)$. A linear Nyquist plot of $\det(I - \hat{T}^{-1}M)$ appears in Figure 3.7-12. It makes no encirclements of the origin (the critical point here is 0, not -1), so the system with cone center is stable. This agrees with the SISO test of Figure 3.7-11, as it should.

To complete the conic sector analysis, compute $\tilde{M} = M(I - \hat{T}^{-1}M)^{-1}$ and then find $\mu = \inf_D \bar{\sigma}(D\tilde{M}D^{-1})$ at each frequency. In this (rather atypical) example, $\tilde{M}$ is already balanced, and the optimal D turns out to be simply the identity matrix I at every frequency. Now set $A = \mu^{1/2}D^{-1}$ and $B = \mu^{1/2}D$; in this case, $A = B = I\bar{\sigma}^{1/2}(\tilde{M})$. Compute $R = \mathrm{diag}(R_i)$, where $R_i = \mathrm{Rad}(\alpha_i, \beta_i)$, and finally plot $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ over frequency. Figure 3.7-13 shows that this stability criterion reaches a maximum amplitude of nearly 3, so this system fails the sufficient condition for stability.

This accurately reflects the fact that the system is unstable for some values of "skew" or timing offset between the two samplers: we have seen that it is unstable when the skew is zero. Since the conic sector method implicitly includes all possible values of skew, from 0 to T, this is as it should be. However, if the system were actually built to have a timing skew of $T/2$, which never changed or drifted, the conic sector test would fail to recognize that the system is stable; in other words, the conic sector method would be conservative.

Two undesirable features of this system combine to make the stability criterion greater than 1. First, the rolloff of $\tilde{M}$ is somewhat slow at high frequencies; it rolls off as $1/\omega^2$. Second, and more importantly, the system with cone center is very close to instability. The Nyquist locus of the loop gain with both samplers replaced by $1/T$ comes very close to the critical point -1, corresponding to a phase margin of only 4.2°. This produces a very large peak in some elements of $\tilde{M}$ near the unity-gain crossover frequency, and this in turn causes the radius R to have large peaks at that frequency and all its aliases. This exhibits a general rule of conic

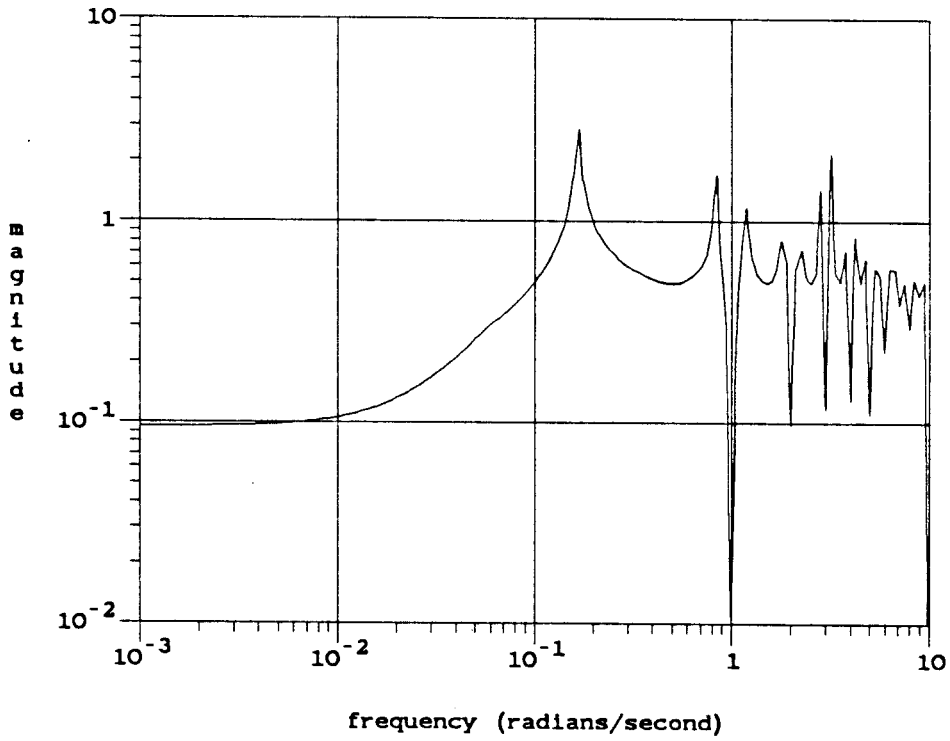

Fig. 3.7-13: Bode plot of $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$. Maximum amplitude is 3, so the system fails the conic sector test for stability.

sector analysis: systems which are close to instability (i.e., with large resonances near the crossover frequency) have large optimal radii, and are more likely to fail the stability test.

It can be argued, then, that conservativeness due to skewed sampling effects, like the forms of conservativeness discussed in the two previous examples, occurs primarily in systems with poor robustness margins — systems which are close to instability anyway. Conversely, it might be expected that systems with good robustness margins will produce little conservativeness in the conic sector method. Preliminary tests on various examples tend to support this conclusion, with one exception: systems with samplers at the outputs of zero-order holds.

3.7.2.4 Sampling the Output of a ZOH

The next example identifies a class of multirate systems for which, in contrast to the previous three examples, the conic sector method as outlined so far gives conservative results even when the systems have good robustness margins. These are systems in which an analog signal is sampled, fed into a zero-order hold, then immediately sampled again at a different rate. If the rates have integer ratios, the problem can be sidestepped successfully by Kranc operator techniques. If the rates have a non-integer ratio, the method of Section 3.7.1 fails. A possible solution to this problem is presented in Section 3.8.

This example is motivated by the author's experiences with switching power supplies; more variations on this same example are explored in Chapter 5. These power converters typically have a lightly damped, second-order lowpass open-loop response. Feedback controllers are used to hold the output voltage constant and equal to a reference input, while rejecting disturbances on the input power line. Figure 3.7-14 shows the system with an analog controller. The plant $G(s)$ has an open-loop transfer function

$$G(s) = \frac{1}{s^2 + .4s + 1}$$

which is a lowpass filter with damping factor $\zeta = 0.2$. Typical power converters have much higher resonant frequencies — about 3000 rad/sec — but this example is scaled down to a resonant frequency of 1. A typical analog controller is

$$K(s) = \frac{312(s+1)^2}{s(s+30)}$$

which has an integrating response at low frequencies, and two zeros at the plant's resonant frequency which allow the loop bandwidth to be extended to 10 radians/sec while maintaining stability. High loop bandwidth is desired in order to give good disturbance rejection at all frequencies. The response to step reference inputs

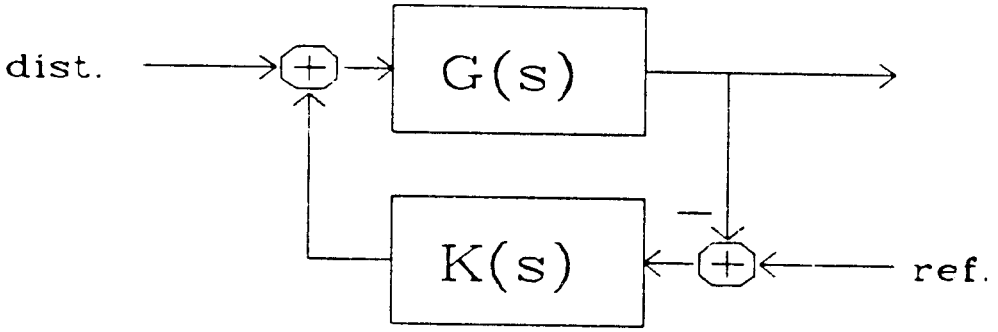

Fig. 3.7-14: An analog feedback loop. $G(s) = 1/(s^2 + .4s + 1)$ and $K(s) = (312(s+1)^2)/(s(s+30))$.

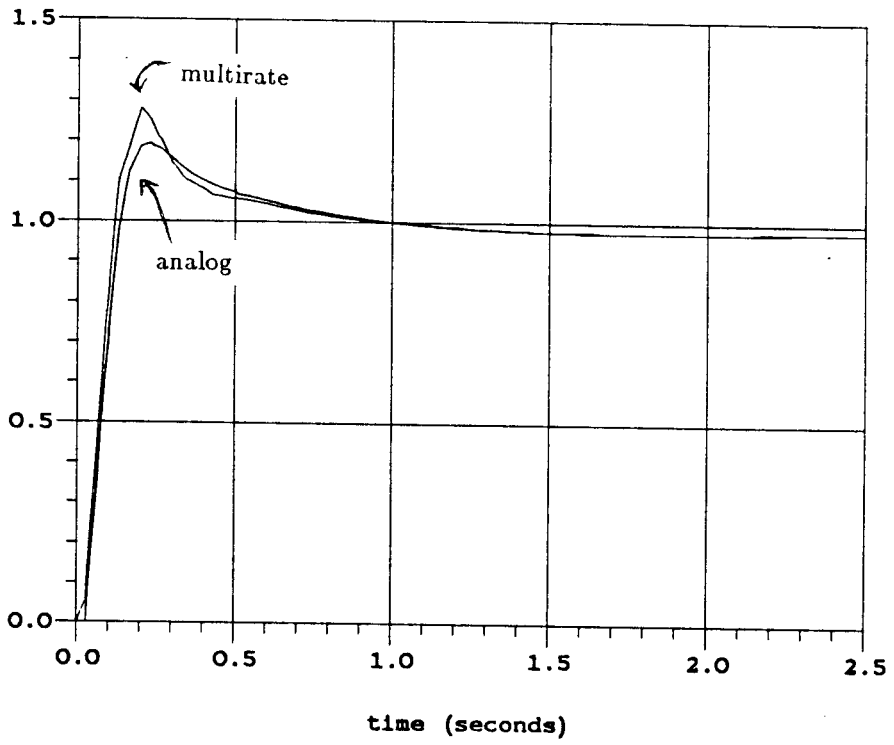

Fig. 3.7-15: Step response to reference input for analog and multirate systems.

(Figure 3.7-15) shows a fairly large 20% overshoot, but this is acceptable since the reference input typically never changes.

Now replace $K(s)$ with the multirate digital controller shown in Figure 3.7-16. To prevent high-frequency switching noise from entering the control loop, an anti-aliasing filter

$$F(s) = \frac{3600}{s^2 + 60s + 3600}$$

is placed between the output of $G(s)$ and the sampler. Suppose the filtered output of $F(s)$ is available at a sample period $T/2$, where $T = 0.1$ second, but it is desired to update the controller's output at period $T/3$. The $T/2$ signal can be held in a buffer register and read at the faster $T/3$ period. This is modelled by feeding the output of the $T/2$ sampler into a ZOH, then sampling the ZOH output at period $T/3$. A digital filter $K_d(z)$ was synthesized, using the complex cepstrum matching technique of Chapter 6, to provide closed-loop response for this system that is similar to that of the original analog system. The philosophy behind this will be explained later; for now, simply let

$$K_d(z) = \frac{491.4(z - .075556)(z - .96736)^2}{(z - 1)(z + .76425)(z - .19620)}$$

where $K_d(z)$ runs at sample period $T/3$.

It is straightforward to use state-space Kranc operator methods, as described in Section 2.8, to determine the nominal stability and dynamic response of this multirate system. Figure 3.7-15 shows the step response, for the reference input, of both the original analog system and this multirate system. They are similar, except that the multirate system has greater overshoot.

Now let's attempt to use the conic sector method of Section 3.7.1 to determine stability for this system. Extract both samplers from the system and place them in the usual diagonal arrangement, with the $T/2$ sampler on top. The interconnection

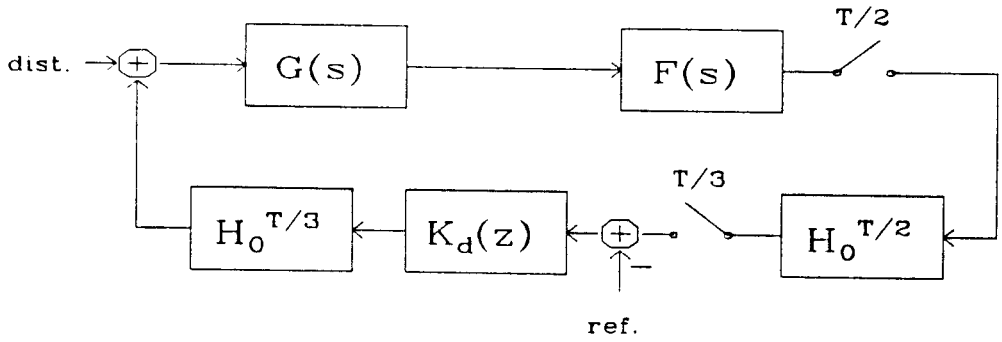

Fig. 3.7-16: Multirate version of Fig. 3.7-14.

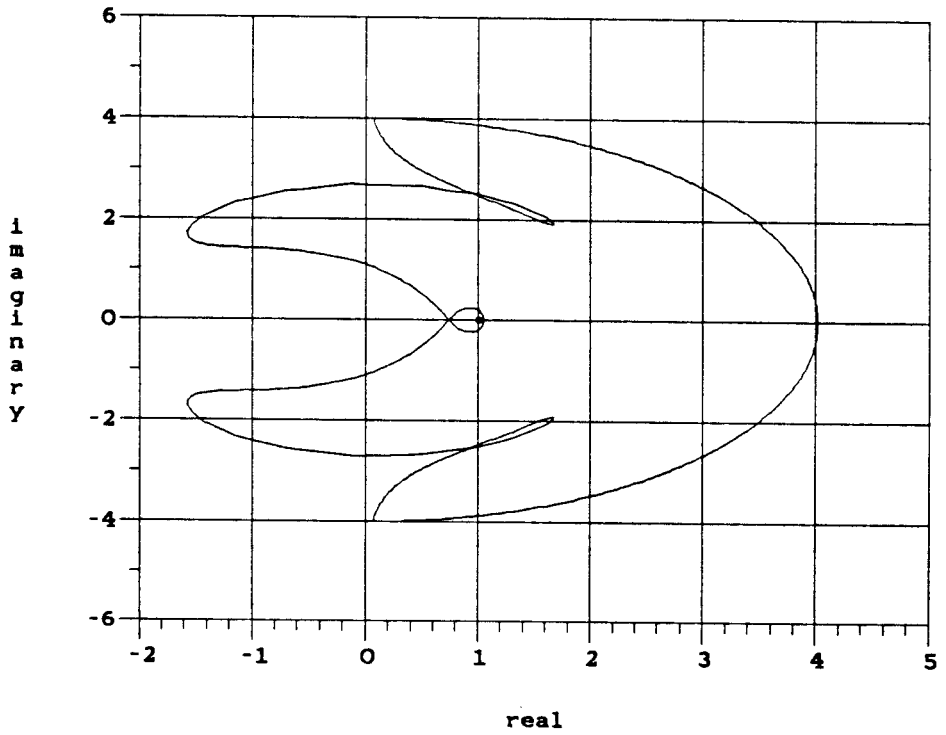

Fig. 3.7-17: Log Nyquist plot of $\det(I - \hat{T}^{-1}M)$. There are no encirclements of 0, so the system is stable with cone center.

matrix M and the matrix $\hat{T}$ are given by

$$M = \begin{bmatrix} 0 & -FGH_0^{T/3}K_d \\ H_0^{T/2} & 0 \end{bmatrix} \qquad \hat{T} = \begin{bmatrix} T/2 & 0 \\ 0 & T/3 \end{bmatrix}$$

The first step is to determine the stability of the system with cone center by plotting $\det(I - \hat{T}^{-1}M)$ as s ranges over a Nyquist D-contour. Since this system has a pole at the origin, the Nyquist contour has a semicircular "notch" there. Figure 3.7-17 shows the log Nyquist plot when the notch has radius .001. The tulip-shaped locus makes no encirclements of the origin, so the system is stable with cone center.

The next step is to compute $\tilde{M} = M(I - \hat{T}^{-1}M)^{-1}$ and $\mu = \inf_D \bar{\sigma}(D\tilde{M}D^{-1})$. Then construct $A = \text{diag}(\alpha_1, \alpha_2) = \mu^{1/2}D^{-1}$ and $B = \text{diag}(\beta_1, \beta_2) = \mu^{1/2}D$ where D is scaled, as always, to satisfy $\bar{\sigma}(D)\underline{\sigma}(D) = 1$. In the 2×2 case, it always happens that $\alpha_1 = \beta_2$ and $\alpha_2 = \beta_1$; this is because $d_1 = 1/d_2$ in the 2×2 case, where $D = \text{diag}(d_1, d_2)$. For this reason, only α_1 and β_1 are plotted in Figure 3.7-18. α_1 is seen to roll off quickly, approximately as $1/\omega^{2.5}$, but β_1 has a disturbing characteristic: it rolls off only as $1/\omega^{1/2}$ as $\omega \to \infty$. This very slow rate of rolloff means that the cone radius formula *does not converge*. Recall that each radius R_i is given by

$$|R_i(j\omega)|^2 = \frac{1}{T_i^2}\left(\sum_n \left|\alpha_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\right|^2\right)\left(\sum_k \left|\beta_i\left(j\omega - jk\frac{2\pi}{T_i}\right)\right|^2\right) - \frac{1}{T_i^2}\sum_n \left|\alpha_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\beta_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\right|^2$$

If $|\beta_i(j\omega)|$ is approximated by $|\beta_i(j\omega)| \approx 1/\omega^{1/2}$, then at low frequencies the radius is approximated by

$$|R_i(j\omega)|^2 \approx \frac{1}{T_i^2}\left(\sum_n \left|\alpha_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\right|^2\right)\left(|\beta_i(j\omega)|^2 + \sum_{k\neq 0}\left|\frac{T_i}{2\pi k}\right|\right) - \frac{1}{T_i^2}\sum_n \left|\alpha_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\beta_i\left(j\omega - jn\frac{2\pi}{T_i}\right)\right|^2$$

and the second infinite series does not converge. Apparently the method of Section 3.7.1 leads to an infinite radius in this example.

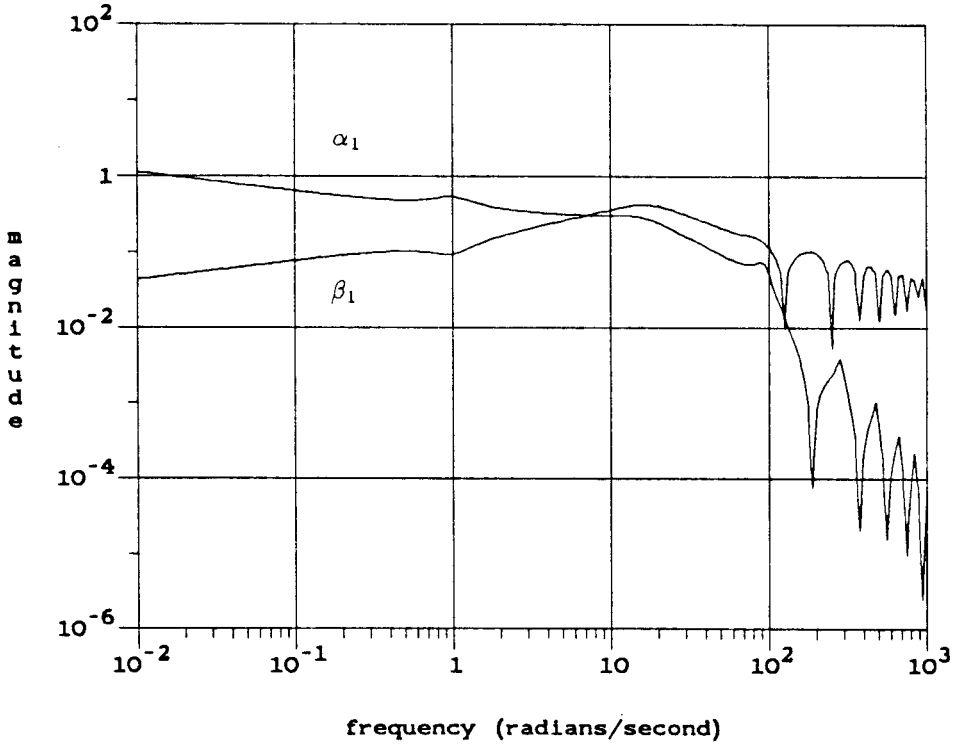

Fig. 3.7-18: Bode plots of α_1 and β_1. β_1 rolls off as $1/\omega^{1/2}$ as $\omega \to \infty$, so the cone radius infinite series does not converge.

The problem is caused by the slow rolloff of the $H_0^{T/2}$ element in M. This type of element will appear in M whenever this type of multirate sampling scheme is modelled by a sampler at the output of a ZOH.

The key to avoiding this problem is to modify the method of Section 3.7.1 slightly. Instead of extracting *all* the samplers in a system into a diagonal arrangement, we should extract only those samplers *which operate on analog signals*. If a sampler operates on a digital signal, which could be the output of another sampler or the output of a digital filter, then it sould be modelled by some other means. When the ratio of sampling rates is an integer, Kranc operator techniques provide a neat solution to this problem.

When the ratio is not an integer, a possible way out of the dilemma is to put *both* samplers and the ZOH inside a *single* conic sector. Theorem 1 must be extended to handle this situation, since it is formulated for a single sampler. The next section derives a modified version of Theorem 1 which handles this case.

3.8 A Conic Sector Containing Two Samplers

Consider the SISO system in Figure 3.8-1. A signal $u(t)$ is filtered by $F(s)$, sampled at period T_1, fed through a hold element $H(s)$, sampled again at period T_2, and finally filtered by $G(s)$ to become the output $y(t)$. The signals at various internal points of the system are labelled x_1 through x_4. This system is linear, though not time-invariant; therefore we can predict its output for any input in L_2 if we know its output for all complex exponential inputs $e^{j\omega t}$. In the following derivation, let $\omega_1 = 2\pi/T_1$ and $\omega_2 = 2\pi/T_2$.

Suppose $x_1 = e^{j\omega t}$. Then the output of the first sampler is

$$x_2 = \frac{1}{T_1}\sum_k e^{j(\omega+k\omega_1)t}$$

The output of the hold element is

$$x_3 = \frac{1}{T_1}\sum_k H(j\omega + jk\omega_1)e^{j(\omega+k\omega_1)t}$$

Sampling again at period T_2 replaces each exponential term $e^{j(\omega+k\omega_1)t}$ with another infinite series:

$$x_4 = \frac{1}{T_1}\sum_k H(j\omega + jk\omega_1)\frac{1}{T_2}\sum_n e^{j(\omega+k\omega_1+n\omega_2)t}$$

Finally, x_4 is filtered by $G(s)$ to produce the output y:

$$y(t) = \frac{1}{T_1T_2}\sum_k\sum_n G(j\omega + jk\omega_1 + jn\omega_2)H(j\omega + jk\omega_1)e^{j(\omega+k\omega_1+n\omega_2)t}$$

Now let $u(t)$ be any signal in L_2; u has a Fourier transform representation

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} U(j\omega)e^{j\omega t}d\omega$$

therefore x_1 is given by

$$x_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)U(j\omega)e^{j\omega t}d\omega$$

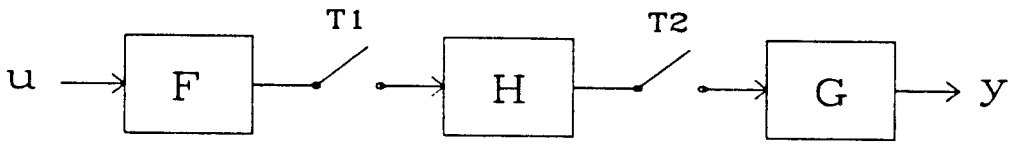

Fig. 3.8-1: A SISO multirate system.

Now by invoking the linearity of the system, we can find the output y with this input u:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{T_1T_2}\sum_k\sum_n G(j\omega+jk\omega_1+jn\omega_2)H(j\omega+jk\omega_1) \quad \times$$

$$F(j\omega)U(j\omega)e^{j(\omega+k\omega_1+n\omega_2)t}d\omega$$

By making a change of variables, this becomes

$$y(t) = \frac{1}{2\pi}\frac{1}{T_1T_2}\int_{-\infty}^{\infty}\Big[\sum_k\sum_n G(j\omega)H(j\omega-jn\omega_2)F(j\omega-jk\omega_1-jn\omega_2) \quad \times$$

$$U(j\omega-jk\omega_1-jn\omega_2)\Big]e^{j\omega t}d\omega$$

Given this input-output description of the system, we may proceed to derive a conic sector which strictly contains it. The derivation closely parallels that of Theorem 1, where the system contains a single sampler.

Let the cone center by given by the baseband (non-aliased) response of the system K:

$$C = \frac{1}{T_1T_2}GHF$$

In order to simplify the notation, define

$$K_{k,n}(j\omega) = \begin{cases} 0 & \text{if } k=n=0; \\ \frac{1}{T_1T_2}G(j\omega)H(j\omega-jn\omega_2)F(j\omega-jk\omega_1-jn\omega_2) & \text{otherwise.} \end{cases}$$

Note that for an input $u(t)$ to this system,

$$y(t)-(Cu)(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_k\sum_n\left(K_{k,n}(j\omega)U(j\omega-jk\omega_1-jn\omega_2)\right)e^{j\omega t}d\omega$$

In order to assure Lebesgue Dominated Convergence, which is required in one step of the derivation, assume for $|\omega|$ sufficiently large that

$$\sum_k\sum_n|K_{k,n}(j\omega)|^2 \le \frac{\alpha}{|\omega|^{1+\beta}} \quad \text{for some } \alpha,\ \beta>0$$

Now we may derive the following Lemma, which is analogous to Lemma 3.A of [20].

Lemma. Let $e(s)$ be the Laplace transform of $e(t) \in L_2$, and let $K_{k,n}$ be as described above. Then

$$\begin{aligned} & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\sum_k\sum_n K_{k,n}(j\omega)e(j\omega - jk\omega_1 - jn\omega_2)\right|^2 d\omega \\ \leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_l\sum_m\sum_k\sum_n |K_{k,n}(j\omega + jl\omega_1 + jm\omega_2)|^2\right)|e(j\omega)|^2 d\omega \end{aligned}$$

Proof:

$$\begin{aligned} & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\sum_k\sum_n K_{k,n}(j\omega)e(j\omega - jk\omega_1 - jn\omega_2)\right|^2 d\omega \\ \leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_k\sum_n |K_{k,n}(j\omega)e(j\omega - jk\omega_1 - jn\omega_2)|^2\, d\omega \\ =\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_k\sum_n |K_{k,n}(j\omega)|^2\,|e(j\omega - jk\omega_1 - jn\omega_2)|^2\, d\omega \\ \leq\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_k\sum_n \left(|K_{k,n}(j\omega)|^2\right)\left(\sum_l\sum_m |e(j\omega - jl\omega_1 - jm\omega_2)|^2\right) d\omega \end{aligned}$$

Now move two of the summation signs outside the integral by Lebesgue Dominated Convergence:

$$= \frac{1}{2\pi}\sum_l\sum_m\int_{-\infty}^{\infty}\left(\sum_k\sum_n |K_{k,n}(j\omega)|^2\right)|e(j\omega - jl\omega_1 - jm\omega_2)|^2\, d\omega$$

Next make a change of variables:

$$\begin{aligned} =\ & \frac{1}{2\pi}\sum_l\sum_m\int_{-\infty}^{\infty}\left(\sum_k\sum_n |K_{k,n}(j\omega + jl\omega_1 + jm\omega_2)|^2\right)|e(j\omega)|^2\, d\omega \\ =\ & \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_l\sum_m\sum_k\sum_n |K_{k,n}(j\omega + jl\omega_1 + jm\omega_2)|^2\right)|e(j\omega)|^2 d\omega \end{aligned}$$

This completes the proof.

Now we may proceed to prove the main theorem. The proof is analogous to the proof of Theorem 1.

Theorem 3. Let K be the two-sampler hybrid operator of Figure 3.8-1. Assume that G, H, and F are stable, and let R and R^{-1} be LTI L_{2e}-stable, causal operators. Let the cone center C be given by $C = \frac{1}{T_1 T_2} GHF$. Then K is strictly inside Cone(C, R) if

$$|R(j\omega)|^2 \quad \geq \quad \frac{1}{1-\epsilon} \sum_l \sum_m \sum_k \sum_n |K_{k,n}(j\omega + jl\omega_1 + jm\omega_2)|^2$$

for all $\omega \in R$ and for some $1 > \epsilon > 0$.

Proof: define the truncated function

$$\tilde{e}_\tau(t) \;=\; \begin{cases} (Re)(t) & \text{if } t \geq \tau; \\ 0 & \text{if } t < \tau. \end{cases}$$

then

$$\begin{aligned}
& \|(K-C)e\|_\tau^2 \\
= \;& \|(K-C)R^{-1}Re\|_\tau^2 \\
= \;& \|(K-C)R^{-1}\tilde{e}_\tau\|_\tau^2 \\
\leq \;& \|(K-C)R^{-1}\tilde{e}_\tau\|_{L_2}^2 \\
= \;& \int_0^\infty \left|(K-C)R^{-1}\tilde{e}_\tau(t)\right|^2 dt \\
= \;& \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\left((K-C)R^{-1}\tilde{e}_\tau\right)(j\omega)\right|^2 d\omega \quad \text{(Parseval's theorem)} \\
= \;& \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\sum_k \sum_n K_{k,n}(j\omega)R^{-1}(j\omega - jk\omega_1 - jn\omega_2)\tilde{e}_\tau(j\omega - jk\omega_1 - jn\omega_2)\right|^2 d\omega
\end{aligned}$$

Now applying the above Lemma:

$$\begin{aligned}
\leq \;& \frac{1}{2\pi}\int_{-\infty}^{\infty} \left(\sum_l \sum_m \sum_k \sum_n |K_{k,n}(j\omega + jl\omega_1 + jm\omega_2)|^2\right) \left|R^{-1}(j\omega)\tilde{e}_\tau(j\omega)\right|^2 d\omega \\
\leq \;& \frac{1}{2\pi}\int_{-\infty}^{\infty} (1-\epsilon)\,|R(j\omega)|^2 \left|R^{-1}(j\omega)\right|^2 |\tilde{e}_\tau(j\omega)|^2 d\omega \\
= \;& \frac{1}{2\pi}\int_{-\infty}^{\infty} |\tilde{e}_\tau(j\omega)|^2 d\omega - \epsilon' \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{|R(j\omega)|^2}{\inf\limits_{\omega \in R} |R(j\omega)|^2} \left|R^{-1}\tilde{e}_\tau(j\omega)\right|^2 d\omega
\end{aligned}$$

where $\epsilon' = \epsilon \left(\inf_{\omega \in R} |R(j\omega)|^2 \right) > 0.$

$$\leq \|\tilde{e}_\tau\|_{L_2}^2 - \epsilon' \|R^{-1}\tilde{e}_\tau\|_{L_2}^2$$

$$\leq \|\tilde{e}_\tau\|_\tau^2 - \epsilon' \|R^{-1}\tilde{e}_\tau\|_\tau^2$$

and by causality of R^{-1}:

$$= \|Re\|_\tau^2 - \epsilon' \|R^{-1}Re\|_\tau^2$$

$$= \|Re\|_\tau^2 - \epsilon' \|e\|_\tau^2$$

$$\leq \|Re\|_\tau^2 - \epsilon'' \left(\|e\|_\tau^2 + \|y\|_\tau^2 \right)$$

where

$$\epsilon'' = \frac{\epsilon'}{1+\alpha^2} > 0 \quad \text{and} \quad \alpha = \sup_{e,\tau} \left(\frac{\|Ke\|_\tau}{\|e\|_\tau} \right) < \infty \quad \text{since } K \text{ is stable.}$$

So the system K meets the definition of being strictly inside $\text{Cone}(C, R)$. This completes the proof.

3.9 Robustness Analysis Using New Results

Now we are ready to put the preceding results together to produce a unified analysis approach for sampled-data systems that is analogous to the LTI approach of Section 2.6, which was built around the Structured Singular Value μ. As in that section, four tests can be made: nominal stability, robust stability, nominal performance, and robust performance. Since the conic sector stability criterion uses a singular value test, we can no longer use μ directly, but in its place we compute $\hat{\mu}(M) = \inf_{D \in \mathcal{D}} \bar{\sigma}(DMD^{-1})$, which is an upper bound for μ. This is not a great sacrifice, because no method is yet known to compute μ exactly for more than three Δ-blocks; therefore $\hat{\mu}$ is generally used instead of μ for LTI analysis as well.

3.9.1 Nominal Stability

Sections 3.4 and 3.7 have shown how to use conic sector theory to determine if a sampled-data feedback system is nominally stable. Section 3.7.1 lists an eight-step procedure for testing nominal stability of MIMO systems. In the SISO case, this procedure gives an optimally small cone radius, as was proved in Section 3.4.1.

3.9.2 Robust Stability

Systems with plant uncertainty can be arranged, as in Figure 3.9-1, with all samplers extracted above the nominal plant M, and with all uncertainty blocks extracted below M in a diagonal arrangement. As in the LTI case, each Δ-block can always be made to have unity magnitude: $\bar{\sigma}(\Delta_i) \leq 1 \ \forall \ \omega$. This is done by lumping any frequency weights together with the plant M.

As with nominal stability analysis, construct $\hat{T} = \text{diag}(T_1, \ldots, T_n)$ where T_i is the period of the ith sampler. Lump $\hat{T}^{-1}$ together with M to form $\tilde{M}$, the operator

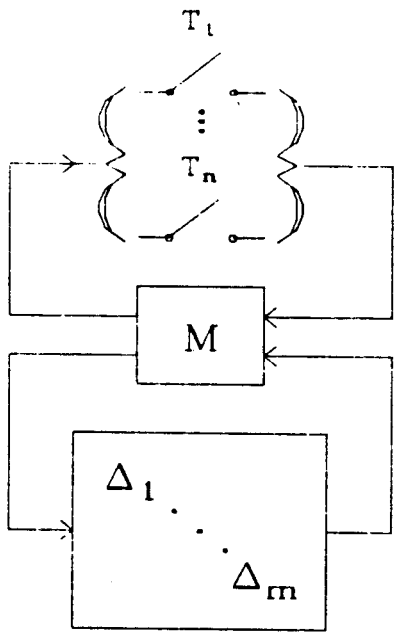

Fig. 3.9-1: A MIMO sampled-data system with structured plant uncertainty.

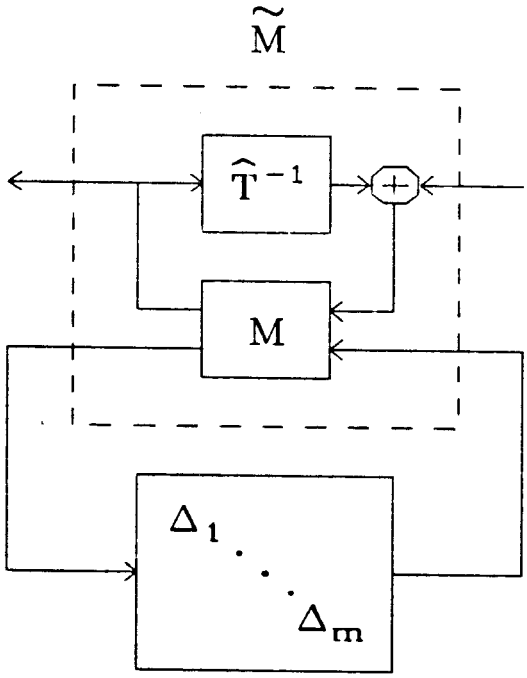

Fig. 3.9-2: The construction of $\tilde{M}$ for the robust stability test, with $\hat{T}^{-1}$ "folded" into M.

in the dashed box of Figure 3.9-2. As before, this has the effect of subtracting the cone centers from the cone-bounded "sandwiched" samplers, leaving them with cone center 0.

The next step is to choose two diagonal operators $A = \text{diag}(\alpha_1, \ldots, \alpha_n, I)$, $B = \text{diag}(\beta_1, \ldots, \beta_n, I)$, and to compute individual cone radii $R_i = \text{Rad}(\alpha_i, \beta_i)$ for the samplers. Each R_i may be factored into $R_i^{1/2} R_i^{1/2}$ as before, and since they are periodic, one $R_i^{1/2}$ may be commuted with each cone-bounded sampler.

This process is identical to the nominal stability procedure, except for the presence of the LTI Δ-blocks. Since the next step requires us to place all the external blocks into a single, diagonal cone-bounded operator, we need a cone bound for each Δ-block. This is easy: since we have assumed $\|\Delta_i\|_\infty \leq 1 \; \forall \; i$, Section 2.2 tells us that

$$\|\Delta_i u\|_{L_2} \;\leq\; \|\Delta_i\|_\infty \|u\|_{L_2} \;\leq\; \|u\|_{L_2} \;\; \forall \, u \in L_2.$$

This implies that each Δ_i is inside $\text{Cone}(0, I)$.

Now define $R^{1/2} = \text{diag}(R_1^{1/2}, \ldots, R_n^{1/2}, I)$ and transfer the LTI Δ-blocks to the upper side of $\tilde{M}$ in Figure 3.9-3. The PLTV operators $\tilde{\Delta}_i$ are zero-center, unity-radius cone-bounded operators obtained, as in Section 3.4.2, by factoring R_i from the input of each $\text{Cone}(0, R_i)$. That is, if $\hat{\Delta}_i R_i$ is inside $\text{Cone}(0, R_i)$, then $\hat{\Delta}_i$ is inside $\text{Cone}(0, I)$. In this way, every PLTV $\hat{\Delta}_i$ block and every LTI Δ_i block is inside $\text{Cone}(0, I)$, so we can use the lemma of Section 3.5 to lump them all into a single diagonal operator K which is inside $\text{Cone}(0, I)$.

The conic sector stability criterion says that this system is stable if

(1) $R^{1/2} A^{-1} \tilde{M} B^{-1} R^{1/2}$ is stable; and

(2) $\|R^{1/2} A^{-1} \tilde{M} B^{-1} R^{1/2}\|_\infty \leq 1$.

Since $R^{1/2}$, A^{-1}, and B^{-1} are stable by construction (they are defined as stable, minimum-phase transfer functions of given magnitude), we need only check the stability of $\tilde{M}$. This can be done by a MIMO Nyquist test on $\det(I - \hat{T}^{-1} M_{11})$.

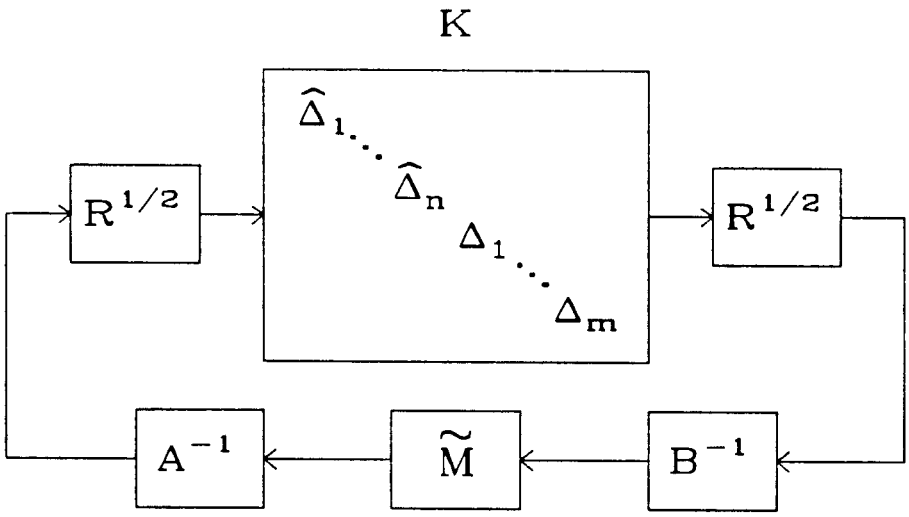

Fig. 3.9-3: After subtracting the cone centers and factoring out R, the PLTV operators $\hat{\Delta}_i$ are each inside Cone$(0, I)$. Lumped together with the LTI operators Δ_i, they form the diagonal operator K, which is inside Cone$(0, I)$.

Before computing $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$ at each frequency, we can exploit one additional trick to reduce the singular value. Although the use of diagonal scaling in the LTV $\hat{\Delta}$-blocks is restricted (since $D_i\hat{\Delta}_i D_i^{-1} \neq \hat{\Delta}_i$ unless $D_i(j\omega)$ is periodic in ω with period $2\pi/T_i$), we may use diagonal scaling on the LTI Δ-blocks. Define the set

$$\mathcal{D} \;=\; \{\ \mathrm{diag}(I_n, d_1 I_{k_1}, \ldots, d_m I_{k_m})\}$$

where k_i is the dimension of the ith LTI Δ-block, and $d_i > 0\ \forall\ i$. Now at each frequency find

$$\hat{\mu} \;=\; \inf_{D\in\mathcal{D}} \bar{\sigma}(DR^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}D^{-1})$$

If $\|\hat{\mu}\|_\infty \leq 1$, the system is robustly stable.

We have not yet addressed the question of how to choose A and B for the robust stability test. A straightforward way is to repeat the method used for the nominal stability test, but using $\tilde{M}_{11}$ instead of the entire matrix $\tilde{M}$, where

$$\tilde{M} \;=\; \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix}$$

and where $\tilde{M}_{11} \in C^{n\times n}$. Define

$$\mathcal{D}_1 \;=\; \{\ \mathrm{diag}(d_1, \ldots, d_n) \mid d_i > 0\ \forall\ i\}$$

and compute

$$\hat{\mu}_1 \;=\; \inf_{D_1\in\mathcal{D}_1} \bar{\sigma}(D_1\tilde{M}_{11}D_1^{-1})$$

Then let $A_1 = \hat{\mu}_1^{1/2}D_1^{-1} = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ and $B_1 = \hat{\mu}_1^{1/2}D_1 = \mathrm{diag}\ (\beta_1, \ldots, \beta_n)$ after scaling D_1 so that $\bar{\sigma}(D_1)\underline{\sigma}(D_1) = 1\ \ \forall\ \omega$.

A less conservative method would be to define $\hat{\mu}_1$ as

$$\hat{\mu}_1 \;=\; \inf_{D_1\in\mathcal{D}_1} \sup_{\Delta\in\mathcal{X}_1} \bar{\sigma}\big(D_1 F_l(\tilde{M}, \Delta)D_1^{-1}\big)$$

and then to proceed as above. A way to do this when there is only one sampler (so that D_1 is a scalar) was discussed in Section 3.4.2. With multiple samplers, however, this appears very difficult.

Let's review the steps for the robust stability test:

1. Set up the system as in Figure 3.9-1, with all samplers and LTI Δ-blocks arranged diagonally about an LTI interconnection matrix M.

2. Construct $\hat{T} = \mathrm{diag}(T_1, \ldots, T_n)$ where T_1 is the ith sampler's period, then "fold" $\hat{T}^{-1}$ into M to produce $\tilde{M}$ as in Figure 3.9-2. $\tilde{M}$ is now given by

$$\tilde{M} = \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix} = \begin{bmatrix} M_{11}(I - \hat{T}^{-1}M_{11})^{-1} & (I - M_{11}\hat{T}^{-1})^{-1}M_{12} \\ M_{21}(I - \hat{T}^{-1}M_{11})^{-1} & M_{22} \end{bmatrix}$$

3. Determine the stability of $\tilde{M}$ by performing the MIMO Nyquist test on $\det(I - \hat{T}^{-1}M_{11})$. If $\tilde{M}$ is stable, proceed.

4. At each frequency ω, compute $\hat{\mu}_1 = \inf_{D_1 \in \mathcal{D}_1} \bar{\sigma}(D_1 \tilde{M}_{11} D_1^{-1})$. (*Optional*: if there is only one sampler, compute instead $\hat{\mu}_1 = \sup_{\Delta \in \mathbf{\Delta}_1} |F_l(\tilde{M}, \Delta)|$ by the method of Section 3.4.2.)

5. Scale D_1 so that $\bar{\sigma}(D_1)\underline{\sigma}(D_1) = 1$ at each frequency.

6. Let $A_1 = \hat{\mu}_1^{1/2} D^{-1} = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ and $B_1 = \hat{\mu}_1^{1/2} D_1 = \mathrm{diag}(\beta_1, \ldots, \beta_n)$. Then let $A = \mathrm{diag}(A_1, I)$ and $B = \mathrm{diag}(B_1, I)$.

7. Compute $R^{1/2} = \mathrm{diag}(R_1^{1/2}, \ldots, R_n^{1/2}, I)$, where $R_i = \mathrm{Rad}(\alpha_i, \beta_i)$.

8. Compute and plot

$$\hat{\mu} = \inf_{D \in \mathcal{D}} \bar{\sigma}(D R^{1/2} A^{-1} \tilde{M} B^{-1} R^{1/2} D^{-1}).$$

(Remember that the first n elements of D are all 1.) If $\|\hat{\mu}\|_\infty \leq 1$, the system is robustly stable.

3.9.3 Nominal Performance

In the LTI analysis method of Section 2.6.4, a system is considered to meet its

nominal performance requirements if its nominal gain is less than 1 at all frequencies: $\|M_0\|_\infty \leq 1$. Since the nominal system M_0 may be defined to include frequency weights and a desired "target" response, this approach is general. Since a sampled-data system is LTV, however, it is not enough to put bounds only on its gain for sine-wave inputs, because this ignores aliasing effects. Instead we may define a system to be within its performance requirements if its input-output relation is inside a conic sector of unity radius and zero center:

$$M_0 \text{ inside Cone}(0, I).$$

As in the LTI case, frequency weights may be included at the input and output of M_0 and a desired input-output response may be included in the model. For analysis purposes, the output of M_0 may be considered to be the frequency-weighted difference between its actual and desired outputs.

How can we use the results of the previous sections to test whether M_0 is inside a given cone? The previous analyses have concentrated on testing the *stability* of the system, when parts of M_0 (the sandwiched samplers) are already known to be inside certain cones. We need to turn this around to test the conicity of M_0 when the stability of a system containing M is already known. But since the conic sector stability criterion is only a sufficient condition for stability, not a necessary and sufficient one, the test can be turned around only under special conditions.

Consider Figure 3.9-4, a feedback loop containing two relations $\mathcal{K}$ and $\mathcal{G}$. The conic sector stability criterion tells us that if $\mathcal{K}$ is inside Cone$(0, I)$ and $\mathcal{G}^I$ is outside Cone$(0, I)$, then this system is stable. Equivalently, it is stable if $\mathcal{K}$ and $\mathcal{G}$ are both inside Cone$(0, I)$. What assumptions can be made about $\mathcal{G}$ which will allow us to reverse this, so that stability implies that $\mathcal{K}$ is inside Cone$(0, I)$?

Lemma. If the system in Figure 3.9-4 is stable for all $\mathcal{G}$ inside Cone$(0, I)$, then $\mathcal{K}$ is inside Cone$(0, I)$.

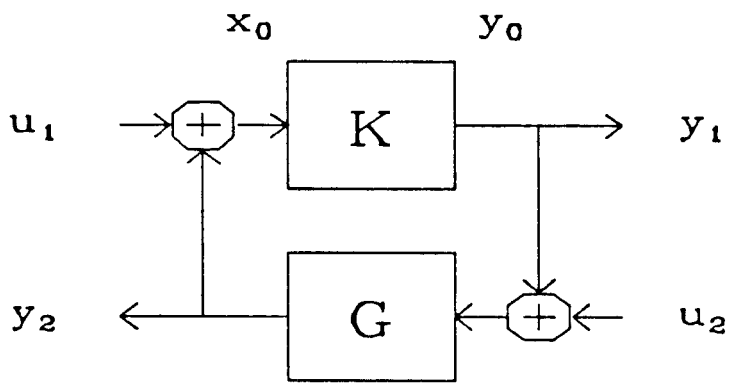

Fig. 3.9-4: A feedback loop containing two relations $\mathcal{K}$ and $\mathcal{G}$.

Proof: Suppose $\mathcal{K}$ is not inside Cone$(0, I)$.

$\Rightarrow$ $\exists\ (x_0, y_0) \in \mathcal{K}$ such that $\|y_0\|_\tau > \|x_0\|_\tau > 0$ for some $\tau > 0$, where $x_0, y_0 \in L_{2e}$

$\Rightarrow$ $\exists\ \mathcal{G}$ inside Cone$(0, I)$ such that $(P_\tau y_0, P_\tau x_0) \in \mathcal{G}$

$\Rightarrow$ $\exists\ y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} P_\tau y_0 \\ P_\tau x_0 \end{pmatrix}$ when $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$ for some $\mathcal{G} \in$ Cone$(0, I)$

$\Rightarrow$ $\|y\|_\tau > k\|u\|_\tau = 0$ for some $\tau \geq 0$, for some $y, u \in L_{2e}$, and $\forall$ finite $k > 0$

Therefore the feedback system is not closed-loop stable for all $\mathcal{G}$ inside Cone$(0, I)$ if $\mathcal{K}$ is not inside Cone$(0, I)$. (See the definition of stability in Section 2.4.) This completes the proof.

Our analysis method, then, is to connect the output of the closed-loop sampled-data system to its input through a "dummy" relation $\hat{\Delta}_{n+1}$, where we allow $\hat{\Delta}_{n+1}$ to range over Cone$(0, I)$, and then to determine stability. If this fictitious system is stable, then the original system is inside Cone$(0, I)$ and the nominal performance requirement is met.

In the previous section, the system was manipulated so that each cone-bounded sampler was modelled by a PLTV operator inside Cone$(0, I)$. All the operators were then combined into a single diagonal operator, also inside Cone$(0, I)$. We can repeat this process, but also include the dummy operator $\hat{\Delta}_{n+1}$. Now the system is reduced to a single, block-diagonal operator inside Cone$(0, I)$ in feedback with an exactly known LTI operator, and the standard stability test may be applied.

Here are the steps for the nominal performance test:

1. Set up the system as in Figure 3.9-5, with all samplers arranged diagonally about an LTI interconnection matrix M.

2. Construct $\hat{T} = \text{diag}(T_1, \ldots, T_n)$ where T_i is the ith sampler's period, then "fold" $\hat{T}^{-1}$ into M to produce $\tilde{M}$ as in Figure 3.9-6. $\tilde{M}$ is now given by

$$\tilde{M} = \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix} = \begin{bmatrix} M_{11}(I - \hat{T}^{-1}M_{11})^{-1} & (I - M_{11}\hat{T}^{-1})^{-1}M_{12} \\ M_{21}(I - \hat{T}^{-1}M_{11})^{-1} & M_{22} \end{bmatrix}$$

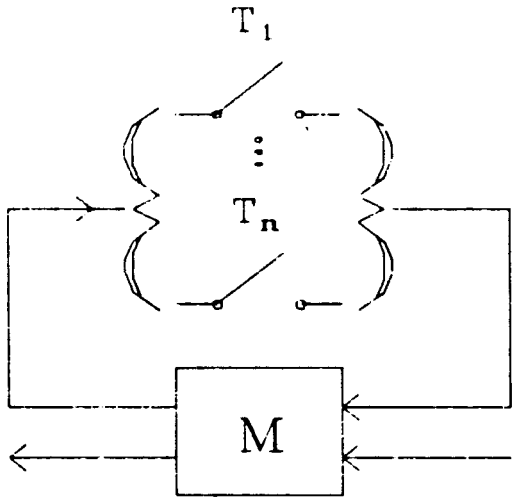

Fig. 3.9-5: A nominal MIMO sampled-data system with input-output response preserved.

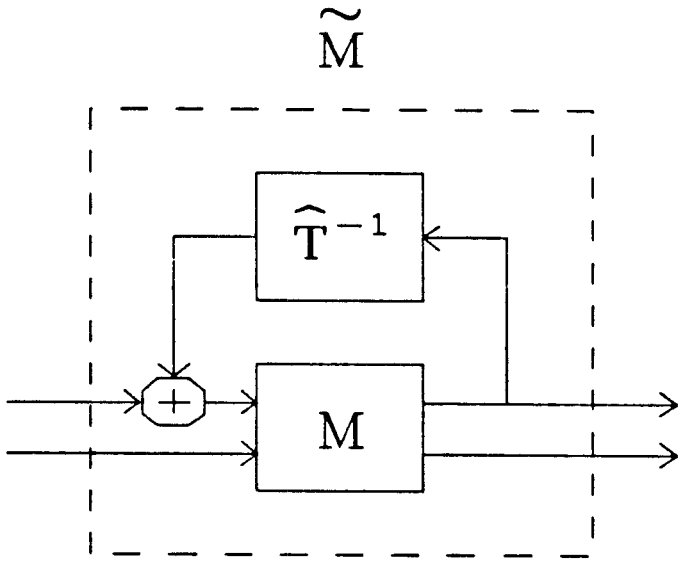

Fig. 3.9-6: The construction of $\tilde{M}$ for the nominal performance test.

3. Determine the stability of $\tilde{M}$ by performing the MIMO Nyquist test on $\det(I - \hat{T}^{-1}M_{11})$. If $\tilde{M}$ is stable, proceed.

4. At each frequency ω, compute $\hat{\mu}_1 = \inf_{D_1 \in \mathcal{D}_1} \bar{\sigma}(D_1 \tilde{M}_{11} D_1^{-1})$, where

$$\mathcal{D} = \{ \text{diag}(d_1, \ldots, d_n) \mid d_i > 0 \, \forall \, i \}$$

5. Scale D_1 so that $\bar{\sigma}(D_1)\underline{\sigma}(D_1) = 1$ at each frequency.

6. Let $A_1 = \hat{\mu}_1^{1/2} D^{-1} = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and $B_1 = \hat{\mu}_1^{1/2} D_1 = \text{diag}(\beta_1, \ldots, \beta_n)$. Then let $A = \text{diag}(A_1, I)$ and $B = \text{diag}(B_1, I)$.

7. Compute $R^{1/2} = \text{diag}(R_1^{1/2}, \ldots, R_n^{1/2}, I)$, where $R_i = \text{Rad}(\alpha_i, \beta_i)$.

8. Compute and plot $\hat{\mu} = \bar{\sigma}(R^{1/2} A^{-1} \tilde{M} B^{-1} R^{1/2})$ over all frequencies. If $\|\hat{\mu}\|_\infty \leq 1$, the system is stable for all dummy blocks $\hat{\Delta}_{n+1}$ inside $\text{Cone}(0, I)$, which implies that the nominal closed-loop system is inside $\text{Cone}(0, I)$. Therefore the system meets its nominal performance requirements.

3.9.4 Robust Performance

By extending the robust performance test of the preceding section to cover a *set* of sampled-data systems, rather than the nominal system only, we get a robust performance test. This test can be defined in a way analogous to the LTI test of Section 2.6.5, except that the performance bound is a conic sector rather than an ∞-norm. That is, a system is considered to meet its robust performance specification if, for all allowed plant perturbations, its input-output relation is inside $\text{Cone}(0, I)$. As in Section 3.9.3, the output used for the test may be the frequency-weighted difference between its actual and desired outputs.

Figure 3.9-7 shows a generic model of a sampled-data system with plant uncertainty. All n samplers are, as usual, extracted from the plant and arranged diagonally above it. In addition, all m LTI uncertainty blocks Δ_i are extracted and arranged diagonally outside the plant. The plant's external input and output are

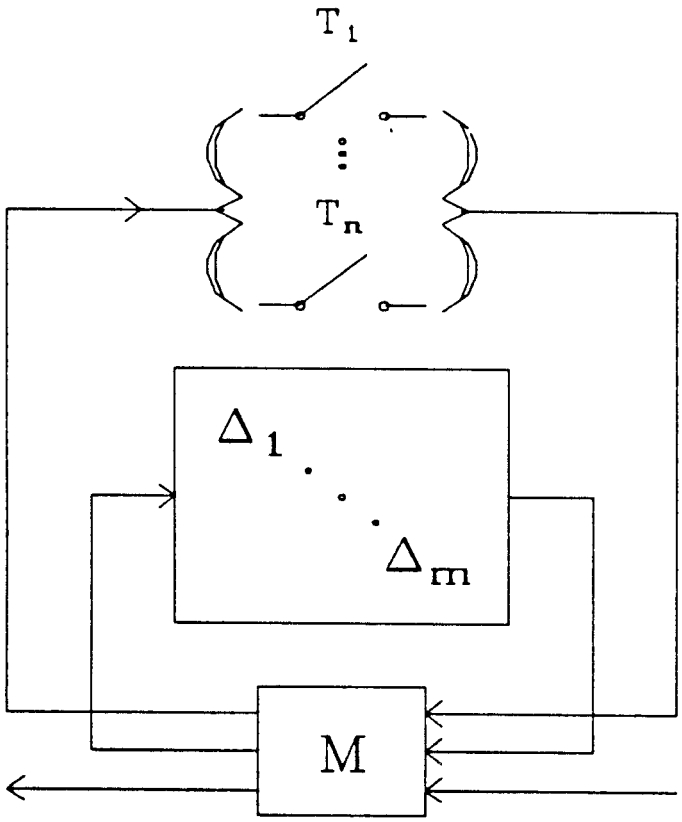

Fig. 3.9-7: An uncertain MIMO system with input-output response preserved.

also shown. The resulting LTI operator M has a 3×3 block structure. Assume that the LTI Δ-blocks are already normalized, so that $\|\Delta_i\|_\infty \leq 1$ for $i = 1$ to m.

We can repeat the procedures of the previous sections to model the LTV effects of each sampler with a cone-bounded PLTV operator $\hat{\Delta}_i$ that is inside Cone$(0, I)$. Then we may determine whether the closed-loop system is inside Cone$(0, I)$ by connecting its output to its input through a "dummy" relation $\hat{\Delta}_{n+1}$, as in Section 3.9.3. If the resulting fictitious feedback system is stable for all $\hat{\Delta}_{n+1}$ inside Cone$(0, I)$, then the robust performance requirement is met.

Here are the steps for the robust performance test:

1. Set up the system as in Figure 3.9-7, with all samplers and LTI Δ-blocks arranged diagonally about an LTI interconnection matrix M.

2. Construct $\hat{T} = \text{diag}(T_1, \ldots, T_n)$ where T_i is the ith sampler's period, then "fold" $\hat{T}^{-1}$ into M to produce $\tilde{M}$ as in Figure 3.9-8. $\tilde{M}$ is now given by

$$\begin{aligned}
\tilde{M} &= \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} & \tilde{M}_{13} \\ \tilde{M}_{21} & \tilde{M}_{22} & \tilde{M}_{23} \\ \tilde{M}_{31} & \tilde{M}_{32} & \tilde{M}_{33} \end{bmatrix} \\
&= \begin{bmatrix} M_{11}(I - \hat{T}^{-1}M_{11})^{-1} & (I - M_{11}\hat{T}^{-1})^{-1}M_{12} & (I - M_{11}\hat{T}^{-1})^{-1}M_{13} \\ M_{21}(I - \hat{T}^{-1}M_{11})^{-1} & M_{22} & M_{23} \\ M_{31}(I - \hat{T}^{-1}M_{11})^{-1} & M_{32} & M_{33} \end{bmatrix}
\end{aligned}$$

3. Determine the stability of $\tilde{M}$ by performing the MIMO Nyquist test on $\det(I - \hat{T}^{-1}M_{11})$. If $\tilde{M}$ is stable, proceed.

4. At each frequency ω, compute $\hat{\mu}_1 = \inf_{D_1 \in \mathcal{D}_1} \bar{\sigma}(D_1 \tilde{M}_{11} D_1^{-1})$, where

$$\mathcal{D}_1 = \{ \text{diag}(d_1, \ldots, d_n) \mid d_i > 0 \; \forall \, i \}$$

5. Scale D_1 so that $\bar{\sigma}(D_1)\underline{\sigma}(D_1) = 1$ at each frequency.

6. Let $A_1 = \hat{\mu}_1^{1/2} D_1^{-1} = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and $B_1 = \hat{\mu}_1^{1/2} D_1 = \text{diag}(\beta_1, \ldots, \beta_n)$. Then let $A = \text{diag}(A_1, I)$ and $B = \text{diag}(B_1, I)$.

7. Compute $R^{1/2} = \text{diag}(R_1^{1/2}, \ldots, R_n^{1/2}, I)$, where $R_i = \text{Rad}(\alpha_i, \beta_i)$.

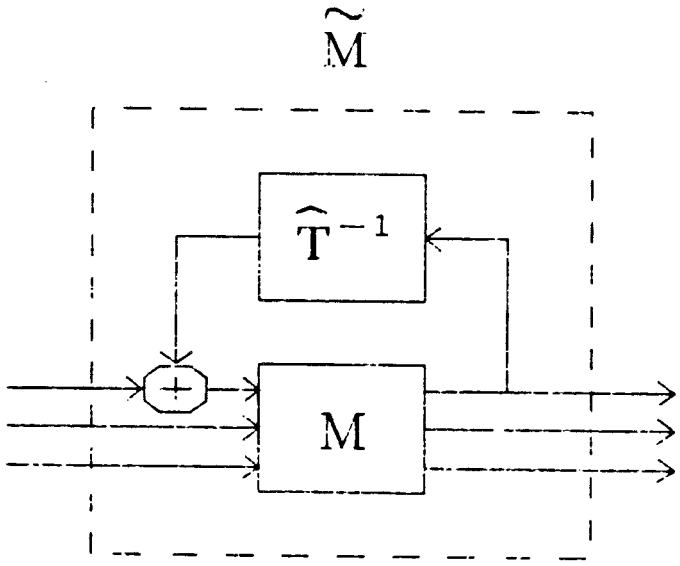

Fig. 3.9-8: The construction of $\tilde{M}$ for the robust performance test.

8. Define

$$\mathcal{D} \;=\; \{\; \mathrm{diag}(I_n, d_1 I_{k_1}, \ldots, d_m I_{k_m}, I) \mid d_i > 0 \;\; \forall\, i \;\}$$

where I_n is an $n \times n$ identity matrix and k_i is the dimension of the ith LTI Δ-block. Over all frequencies, compute and plot

$$\hat{\mu} \;=\; \inf_{D \in \mathcal{D}} \bar{\sigma}(D R^{1/2} A^{-1} \tilde{M} B^{-1} R^{1/2} D^{-1})$$

(Remember that the first n elements of D are all 1.) If $\|\hat{\mu}\|_\infty \leq 1$, the system is stable for all dummy blocks $\hat{\Delta}_{n+1}$ inside $\mathrm{Cone}(0, I)$ and for all LTI Δ-blocks such that $\|\Delta_i\|_\infty \leq 1$, which implies that the closed-loop system is inside $\mathrm{Cone}(0, I)$ for all allowed perturbations. Therefore the system meets its robust performance requirements.

The only significant difference beween this procedure and the nominal performance test of Section 3.9.3 is that diagonal scaling is allowed in the LTI Δ-block channels. Diagonal scaling is not used in the PLTV $\hat{\Delta}_i$ channels, because for an LTI operator D_i, $D_i \hat{\Delta}_i D_i^{-1} \neq \hat{\Delta}_i$ unless $D_i(j\omega)$ is periodic in ω with period $2\pi/T_i$. Such a restriction on D is very difficult to handle algorithmically, and the pre-scaling of $\tilde{M}_{11}$ by D_1 prior to computing the radius should reduce the need for scaling in those channels.

Frequency-independent scaling cannot be used on the "dummy" channel $\hat{\Delta}_{n+1}$ because $\hat{\Delta}_{n+1}$ cannot be assumed to be LTI; it must be allowed to range over all of $\mathrm{Cone}(0, I)$. In fact, it cannot even be assumed to be linear.

Chapter Four — Practical Considerations

4.1 Computing the Cone Center and Radius

4.1.1 State-space Method for SISO Operators

The basic theorem of Section 3.1 states that a SISO hybrid operator with pre-filter $F(s)$, sampling period T, and hold $H(s)$ is inside Cone(C, R) where $C(s) = \frac{1}{T}H(s)F(s)$ and

$$|R(j\omega)|^2 = \frac{1}{T^2}\left\{\left(\sum_k |F_k(j\omega)|^2\right)\left(\sum_k |H_k(j\omega)|^2\right) - \left(\sum_k |F_k(j\omega)H_k(j\omega)|^2\right)\right\}$$

If $F(s)$ and $H(s)$ are proper rational functions, they can be represented by state-space quadruples:

$$F(s) = \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & 0 \end{array}\right] \qquad H(s) = \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & 0 \end{array}\right]$$

It is assumed that the D terms are zero, since the infinite series would not converge if they were not: $F(s)$ and $H(s)$ must roll off at high frequencies. Given these state-space quadruples, we can compute the three infinite series in closed form using only real matrix multiplication and a matrix exponential routine, by using a method presented in [20]. Recall from basic sampling theory that the z-transform of a discretized analog plant, evaluated on the unit circle, is $1/T$ times the infinite summation of the aliased analog transfer function:

$$G^T(z) = \frac{1}{T}\sum_k G_k(j\omega)$$

We saw in Section 2.7.2 that $G^T(z)$ may be computed from $G(s)$ in state space:

$$\text{if} \quad G(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \quad \text{then} \quad G^T(z) = \left[\begin{array}{c|c} e^{AT} & e^{AT}B \\ \hline C & D \end{array}\right]$$

A transfer function which equals $|F(j\omega)|^2$ on the $j\omega$-axis can be constructed in state space by taking advantage of the fact that $|F(j\omega)|^2 = F^*(j\omega)F(j\omega)$ where $F^*(j\omega)$ is the complex conjugate of $F(j\omega)$. (F^* is the adjoint operator of F.) Using the rules in Section 2.2.1 for taking the adjoint of a system and the cascade of two systems, we get

$$\begin{aligned} F^*(s)F(s) &= \left[\begin{array}{c|c} -A_1^T & -C_1^T \\ \hline B_1^T & 0 \end{array}\right] \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & 0 \end{array}\right] \\ &= \left[\begin{array}{cc|c} -A_1^T & -C_1^T C_1 & 0 \\ 0 & A_1 & B_1 \\ \hline B_1^T & 0 & 0 \end{array}\right] \equiv \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & 0 \end{array}\right] \end{aligned}$$

Now find

$$\frac{1}{T}\sum_k |F_k(j\omega)|^2 = (F^*F)^T(z) = \left[\begin{array}{c|c} e^{\hat{A}T} & e^{\hat{A}T}\hat{B} \\ \hline \hat{C} & 0 \end{array}\right]$$

where $z = e^{j\omega T}$.

Repeat this process to find $\frac{1}{T}\sum_k |H_k(j\omega)|^2$ and $\frac{1}{T}\sum_k |F_k(j\omega)H_k(j\omega)|^2$. For the latter summation, first cascade F and H, then compute $(FH)^*(FH)$. In terms of the three summations, the squared cone radius is

$$\begin{aligned} |R(j\omega)|^2 &= \left(\frac{1}{T}\sum_k |F_k(j\omega)|^2\right)\left(\frac{1}{T}\sum_k |H_k(j\omega)|^2\right) - \frac{1}{T}\left(\sum_k |F_k(j\omega)H_k(j\omega)|^2\right) \\ &= (F^*F)^T(z)(H^*H)^T(z) - \frac{1}{T}(H^*F^*FH)^T(z) \end{aligned}$$

where $z = e^{j\omega T}$.

This procedure works if $H(s)$ is a rational function in s. More typically, however, H is a zero-order hold: $H(s) = H_0^T(s) = (1 - e^{-sT})/s$. Since this is an infinite-dimensional system, no state-space quadruple exists for it. However, since the

factor $1 - e^{-j\omega T}$ is periodic in ω, it can be pulled outside the summation:

$$\sum_k |H_k(j\omega)|^2 = \sum_k \left| \frac{1 - e^{-j(\omega T + 2\pi k)}}{j(\omega + 2\pi k/T)} \right|^2$$

$$= \sum_k \left| \frac{1 - e^{-j\omega T}}{j(\omega + 2\pi k/T)} \right|^2 = \left| 1 - e^{-j\omega T} \right|^2 \sum_k |\hat{H}_k(s)|^2$$

where $\hat{H}(s) = 1/s$, a simple integrator.

Therefore we may handle the case of a ZOH by setting $H(s) = \hat{H}(s) = 1/s$ and applying the state-space method described above, then multiplying the resulting value of $|R(j\omega)|^2$ by $|1 - e^{-j\omega T}|^2 = (1 - z)(1 - z^{-1})$:

$$|R(j\omega)|^2 = \left\{ (F^*F)^T(z)(\hat{H}^*\hat{H})^T(z) - \frac{1}{T}(\hat{H}^*F^*F\hat{H})^T(z) \right\} (1 - z)(1 - z^{-1})$$

where $z = e^{j\omega T}$.

This method, although accurate and appealingly simple, is not well suited for the new nonconservative conic sector results of this thesis. This is because the new methods set $F(j\omega) = \alpha(j\omega)$ and $H(j\omega) = \beta(j\omega)$, where α and β are not rational functions in s. Therefore no state-space quadruples exist for α and β. This requires us to approximate the radius numerically by first finding α and β for a large number of frequency points, then computing the truncated infinite series. This technique is described in the next section.

4.1.2 Numerical Method for SISO Operators

The new analysis methods of this thesis require solving numerical minimization problems at many frequency points. For instance, to make the MIMO nominal stability test, we solve

$$\hat{\mu} = \inf_{D \in \mathcal{D}} \bar{\sigma}(D\tilde{M}D^{-1})$$

at each frequency, where $\mathcal{D}$ is a set of positive real diagonal matrices. The next step

is to let

$$A = \hat{\mu}^{1/2} D^{-1} = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$$

$$B = \hat{\mu}^{1/2} D = \mathrm{diag}(\beta_1, \ldots, \beta_n)$$

and then to compute n cone radii $R_i = \mathrm{Rad}(\alpha_i, \beta_i)$ using the SISO version of the basic theorem. Therefore α_i and β_i are typically given as a set of logarithmically spaced data points. The state-space method of Section 4.1.1 is not applicable, since we have no state-space representation of α_i and β_i; in fact, α_i and β_i are usually not rational functions of s.

The formula for the cone radius is

$$|\mathrm{Rad}(\alpha, \beta)|^2 = \frac{1}{T^2} \left\{ \left(\sum_k |\alpha_k(j\omega)|^2 \right) \left(\sum_k |\beta_k(j\omega)|^2 \right) - \left(\sum_k |\alpha_k(j\omega)\beta_k(j\omega)|^2 \right) \right\}$$

The three infinite series can be approximated numerically by truncating each series at a sufficiently high number of terms. Since the input data is available only at certain fixed frequency points, most of the aliased terms $\alpha_k(j\omega)$ and $\beta_k(j\omega)$ must be estimated by interpolating between the points that are available. For large values of k, the aliased frequency may be well above the highest frequency point in the input data. In these cases, $\alpha_k(j\omega)$ and $\beta_k(j\omega)$ may be estimated by assuming that at high frequencies,

$$|\alpha(j\omega)| \leq \frac{K_\alpha}{\omega^a} \quad \text{and} \quad |\beta(j\omega)| \leq \frac{K_\beta}{\omega^b} \quad \text{for some } a, b, K_a, K_b > 0$$

The constants a and b set the rate of rolloff at high frequencies. The faster α and β roll off, the smaller $\mathrm{Rad}(\alpha, \beta)$ becomes.

Most of the cone radius plots in this thesis were generated by a simple FORTRAN program using this approach. The program reads two data files containing α and β at logarithmically spaced points, then prompts the user for the high-frequency rolloff constants a and b. It computes the three truncated infinite series using linear interpolation at low frequencies and the rolloff bounds at high frequencies. Then it

combines the three series by the above formula and writes $\mathrm{Rad}(\alpha, \beta)$ to an output file, for the same frequency points as the input files α and β. Although not as elegant as the state-space method of Section 4.1.1, the program runs quickly and is well suited to the new conic sector methods of this thesis.

It is not necessary to compute cone radii using the MIMO version of the basic theorem, since the methods of this thesis treat MIMO systems by a series of applications of the SISO version of the theorem. That is, each SISO sampler is first placed within its own conic sector, then recombined with others to form a MIMO operator.

4.2 Matrix Balancing and Diagonal Scaling

4.2.1 Osborne's Method

The multivariable results of Chapter 3 depend on computing a diagonal matrix $D \in \mathcal{D}$ which minimizes $\bar{\sigma}(D\tilde{M}D^{-1})$. Since this is not an analytic function of D, the minimization problem is complicated, and available software to solve it converges rather slowly. This problem is discussed further in the next section. A closely related problem with an efficient, elegant solution algorithm is $\inf_{D\in\mathcal{D}} \|D\tilde{M}D^{-1}\|_F$, where $\|\cdot\|_F$ is the Frobenius norm.

This problem can be solved by a rapidly convergent iterative method due to Osborne [13]. Recall that the squared Frobenius norm of a matrix is simply the sum of the squared magnitudes of all its elements; i.e.,

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 \quad \text{for } A \in C^{n\times n}$$

For simplicity, assume that each diagonal element d_j of $D = \mathrm{diag}(d_1, \ldots, d_n)$ is a free variable; when some elements are repeated, the method can be easily modified.

Now observe that when an element d_k is varied, only the kth row and kth

column of DAD^{-1} are changed; furthermore, the kth diagonal element of DAD^{-1} is unchanged. This is because

$$\left(DAD^{-1}\right)_{ij} = \frac{d_i a_{ij}}{d_j}$$

Since $\|DAD^{-1}\|_F^2$ is simply the sum of all squared elements of DAD^{-1}, it may be decomposed into three parts: a term proportional to d_k^2, a term proportional to $1/d_k^2$, and a term independent of d_k^2.

$$\begin{aligned}
\|DAD^{-1}\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{d_i^2|a_{ij}^2|}{d_j^2} \\
&= d_k^2 \sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{|a_{kj}|^2}{d_j^2} + \frac{1}{d_k^2}\sum_{\substack{i=1 \\ i\neq k}}^{n} d_i^2|a_{ij}|^2 + \left(\sum_{\substack{i=1 \\ i\neq k}}^{n}\sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{d_i^2|a_{ij}|^2}{d_j^2} + |a_{kk}|^2\right) \\
&= d_k^2 R + \frac{1}{d_k^2} S + C
\end{aligned}$$

Simple calculus shows that this expression is minimized when

$$d_k^2 = \sqrt{\frac{S}{R}}$$

Therefore the Frobenius norm may be minimized with respect to a single element of D in only one step. Osborne's algorithm to find the optimal D is to let k step from 1 through n, then back to 1, until the norm has converged:

1. Let $k = 1$ and $D = I$.

2. Compute the sums

$$R = \sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{|a_{kj}|^2}{d_j^2} \quad \text{and} \quad S = \sum_{\substack{i=1 \\ i\neq k}}^{n} d_i^2|a_{ik}|^2$$

3. Let $d_k = (S/R)^{1/4}$.

4. Test for convergence. If converged, stop; otherwise increment k by 1. If $k = n$, let $k = 1$ again.

5. Go to Step 2.

This method always converges and usually does so very quickly. Only a small amount of computation is involved, since the sums R and S have only $n-1$ terms each.

Since $\|\cdot\|_F$ and $\|\cdot\|_2$ are compatible norms, the D which minimizes $\|D\tilde{M}D^{-1}\|_F$ usually comes close to minimizing $\|D\tilde{M}D^{-1}\|_2 = \bar{\sigma}(D\tilde{M}D^{-1})$ as well:

$$\|D\tilde{M}D^{-1}\|_2 \;\leq\; \|D\tilde{M}D^{-1}\|_F \;\leq\; n\|D\tilde{M}D^{-1}\|_2$$

For this reason, it is often sufficient simply to apply Osborne's method alone when using the methods of Chapter 3 for MIMO sampled-data system analysis. That is, instead of solving $\hat{\mu} = \inf_{D\in\mathcal{D}} \bar{\sigma}(D\tilde{M}D^{-1})$ directly, solve $\inf_{D\in\mathcal{D}} \|D\tilde{M}D^{-1}\|_F$ and then compute $\hat{\mu} = \bar{\sigma}(D\tilde{M}D^{-1})$ with the resulting D. This can save a considerable amount of computer time without significantly degrading the results.

4.2.2 Singular Value Minimization

For cases where the Osborne method is not enough, other iterative techniques exist to directly minimize $\bar{\sigma}(D\tilde{M}D^{-1})$. Since the singular values of a matrix are not, in general, differentiable functions of the matrix elements, a gradient cannot always be defined for steepest descent methods. In [5], however, Doyle defines a generalized gradient which serves the same purpose of providing a search direction which reduces $\bar{\sigma}(D\tilde{M}D^{-1})$. Since this function can be shown to be convex in D, the function always converges to its global minimum if it is reduced at each iteration. A complete description of the method is beyond the scope of this thesis; see [5] for details.

In Chapter 3, $\inf_{D\in\mathcal{D}} \bar{\sigma}(D\tilde{M}D^{-1})$ is desired for direct use; in analysis of purely LTI systems, it is desired only as an upper bound for the true Structured Singular Value μ. In addition to [5], another article discussing the computation of μ is [8].

4.3 Application to Typical Control System Specifications

This section discusses how to use the conic sector methods of this thesis to test typical control system specifications. These include gain and phase margins, loop-at-a-time margins for multiloop systems, RMS output error for sine-wave and white noise inputs, step response, and rise time.

4.3.1 Gain Margin, Phase Margin, and Gain-phase Margin

For a system with a single feedback loop, gain margin is usually defined as the amount by which the loop gain can be increased or reduced, without phase shift, before the system becomes unstable. Refer to Figure 4.3-1; the system has been drawn as a lower fractional transformation so that its single SISO feedback loop passes through the parameter α, which is nominally 1. Any single-loop system may be redrawn in this way. The gain margin is the interval on the real number line, containing 1, of values for α which stabilize the system. A typical gain margin (GM) is GM $= [0.1, 2] = [-20$ dB, $+6$ dB$]$.

Phase margin (PM) is defined in a similar way, but instead of varying α over the real numbers, we vary it over the unit-magnitude complex numbers: $\alpha = e^{j\phi}$, where $\phi \in R$. The phase margin is the range of real values of ϕ, containing 0, for which the system is stable. The interval is always symmetrical about 0, because changing ϕ has the effect of rotating the Nyquist locus, which is itself symmetrical around the real axis. A typical phase margin is PM $= [-\pi/4, \pi/4] = [-45^o, +45^o]$.

The phase margin is rather unrealistic, since an operator α with equal phase shift at all frequencies is unrealizable. It is useful, however, in showing nearness to instability in a general way, and in giving a rough indication of the nominal system's damping at the unity loop gain crossover frequency.

Another single-loop robustness margin is sometimes called the gain-phase mar-

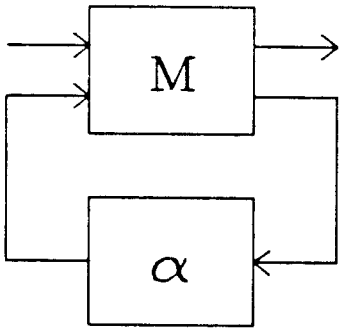

Fig. 4.3-1: Calculation of gain and phase margins. α is nominally 1; it is varied along the real axis or the unit circle until the loop is destabilized.

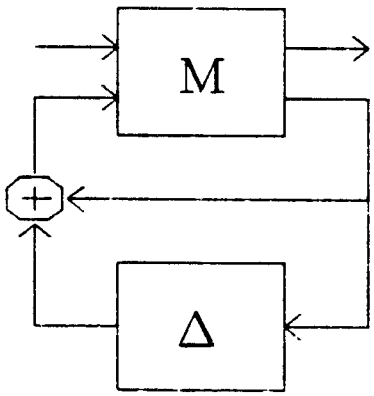

Fig. 4.3-2: Calculation of gain-phase margin. Δ is inside the disk $|\Delta| < \beta$; β is increased until some Δ in the disk destabilizes the loop.

gin. Replace α by $1+\Delta$ where $|\Delta| < \beta$ and Δ ranges over all complex values inside a disk of radius β. The gain-phase margin is the largest value for β which maintains stability for all $|\Delta| < \beta$. The system is redrawn in Figure 4.3-2 with $1+\Delta$ replacing α. If M is LTI, the gain "seen" by α is simply $M_{22}/(1-M_{22})$; so by the small gain theorem, the system is stable if $|\beta M_{22}/(1-M_{22})| < 1$ at all frequencies. Therefore the gain-phase margin is

$$\beta_{\max} = \inf_{\omega \in R} \left| \frac{1 - M_{22}(j\omega)}{M_{22}(j\omega)} \right|$$

The gain-phase margin is well matched to the methods of this thesis, since it models plant uncertainty as a complex disk: $\|\Delta\|_\infty < \beta$. By including the scaling factor β with the plant, we get the usual normalized form $\|\Delta\|_\infty < 1$. For a system with one or more samplers, then, the gain-phase margin can be determined by increasing β from 0 until the system fails the robust stability test $\|\hat{\mu}\|_\infty \leq 1$ as described in Section 3.9.2. Increasing β is equivalent to scaling one of the rows or columns of the interconnection matrix $\tilde{M}$. An example of this procedure is in Section 5.1.

Gain margins and phase margins, however, are less well suited to the Δ-block representation of uncertainty used in this thesis, because each Δ-block is always assumed to take on all the values in a complex disk. In the definition of gain margin, α takes only real values; for phase margin, it takes on values along an arc in the complex plane. Although it is always possible to enclose these loci in disks, it usually leads to conservativeness, since such a disk includes many values for α which are not of interest. With conic sector analysis of single-loop sampled-data systems, then, the gain and phase margins can be computed only with an indirect, iterative method: substitute various values for α along a real line (or an arc) and then make the nominal stability test $\|R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}\|_\infty < 1$. Continue varying α until the system fails the stability test. In this way a guaranteed gain or phase margin may be generated. Since the conic sector method provides only a sufficient

condition for stability, the system may be stable outside the guaranteed interval, so some conservativeness remains.

If a single-loop system has only one sampler or has several synchronized samplers with rates related by rational numbers, we can fall back on LTI methods and use the Kranc operator robustness analysis of Section 2.8.2. This allows us to produce exact (necessary and sufficient) gain, phase, and gain-phase margins for such systems. It will be used in Chapter 5 to test the conservativeness of the new conic sector methods. Kranc robustness analysis involves computing the characteristic loci (eigenvalues) versus frequency for the matrix-valued Kranc operator loop transfer function. For a single-rate system, the matrix becomes a scalar.

It should be noted that gain, phase and gain-phase margins all make unreasonable assumptions about plant uncertainty: they all assume that the nominal plant is perturbed in a way that is the *same at all frequencies*. This is very unrealistic. A more logical way to define system robustness is to use engineering judgment at the beginning of an analysis, and define frequency-dependent models for plant uncertainty at the same time as models for the plant itself. In this way, each plant subsystem is treated from the beginning, not with a single, unrealistic model, but with a *set* of models. If the uncertainty at each frequency can be modelled by an interconnection of disk-bounded Δ-blocks, then the frequency weights may be absorbed into the plant, resulting in the block-diagonal $\|\Delta\|_\infty < 1$ model used throughout this thesis.

4.3.2 Loop-at-a-time Margins

Multi-loop feedback systems are sometimes required to meet loop-at-a-time gain, phase, or gain-phase margin requirements. These are defined by setting the loop gains in all but one of the loops to their nominal values, then varying the

gain in the remaining loop in the manner discussed in the previous section until the system becomes unstable. This approach gives one a qualitative idea of how close a system is to instability, but it can lead to a false sense of security. This is because it ignores the effects of allowing the gains in all loops to vary simultaneously, which is what happens in actual operation. Examples can be constructed [6] of multi-loop systems with excellent loop-at-a-time margins which are arbitrarily close to instability when all their channels are perturbed at once.

The multiple Δ-block model of plant uncertainty due to Doyle [5],[6] and used throughout this thesis inherently includes the effects of perturbations in all channels at once.

4.3.3 RMS Output Error for Sinusoidal Inputs

Feedback regulators, which have the purpose of holding their outputs constant when their inputs are disturbed, are often specified in terms of allowable root-mean-square output error when input sine waves of given frequency and magnitude are applied. If the closed-loop regulator is known to be inside a zero-center conic sector $\mathrm{Cone}(0, R)$, this type of specification is trivially easy to test. Call the input u and the output e, and let $u = u_0 e^{j\omega t}$ where $u_0 \in C^n$; then the RMS output is

$$\|e\|_{\mathrm{RMS}} \equiv \lim_{\tau\to\infty} \sqrt{\frac{1}{\tau}\int_0^\tau e^*(t)e(t)dt} = \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|e\|_\tau \leq \bar{\sigma}(R(j\omega))\|u_0\|_2$$

so the cone radius provides a direct bound on the RMS output. This follows from the definition of a conic sector. Call the regulator's input-output operator K; then if K is inside $\mathrm{Cone}(0, R)$,

$$\|e\|_\tau \leq \|Ru\|_\tau \quad \forall\, \tau > 0$$

As $\tau \to \infty$, the steady-state response of the LTI operator R dominates:

$$\lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|e\|_\tau \le \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|Ru\|_\tau = \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|R(j\omega)u_0 e^{j\omega t}\|_\tau$$

$$\le \bar{\sigma}(R(j\omega))\|u_0\|_2$$

If the cone-bounded operator is SISO, we can drop the pretense that the input is complex-valued. Let $u = k\sin\omega t$, where $k > 0$. Then

$$\|e\|_{\text{RMS}} \le \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|Rk\sin\omega t\|_\tau = k|R(j\omega)| \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|\sin\omega t\|_\tau$$

$$= \frac{k|R(j\omega)|}{\sqrt{2}}$$

Conic sector methods, then, are well matched to this type of specification. We may use the nominal performance or robust performance methods of Sections 3.9.3 and 3.9.4 to find the smallest possible cone radius R such that the closed-loop system is inside Cone$(0, R)$. Recall that these methods determine whether the input-output operator is inside Cone$(0, I)$. We can convert this to Cone$(0, R)$ by adding trial frequency weights R^{-1} to the input of the system, *not the output* , before performing the above test. This is valid because

$$K \text{ inside Cone}(0, R) \quad \Leftrightarrow \quad KR^{-1} \text{ inside Cone}(0, I)$$

Since this statement holds only for R^{-1} at the input of K, not the output, the trial values of R^{-1} must be applied to the input. By increasing R^{-1} until the system fails the Cone$(0, I)$ test, we can find the smallest possible radius.

Because the computation of Rad(α, β) involves the products of many aliased terms, there will generally be interdependence between the magnitudes of R at different frequencies. That is, R can be decreased at one frequency in exchange for increasing it at another. To avoid being mired in such complications, it is best to first work directly from the system specifications to find an R such that, if the system K is inside Cone$(0, R)$, it passes its specifications. Then scale R^{-1} into the plant's input and apply the Cone$(0, I)$ test.

The above description assumes a cone center of 0. In general, when the methods of this thesis are applied, it is more natural to try to derive a conic sector with nonzero center. This can be done by replacing each LTI uncertainty Δ-block with 0, and each sampler with $1/T_i$, where T_i is its period. Then compute the input-output transfer function for the resulting LTI system, and call it C. This provides a good choice for an input-output cone center. C is then subtracted from the appropriate elements of the interconection matrix $\tilde{M}$, and the procedure above may be applied to the modified $\tilde{M}$. Now the tests will determine if the input-output response is inside Cone(C, R). This is especially important for servo systems, for which the desired response is nonzero.

When the system is inside a cone with nonzero center, the remarks made about RMS output must be amended. The system can be conceptually broken into two parallel parts: an exactly known LTI operator C, and a partially unknown operator K inside Cone$(0, R)$. The earlier remarks apply to K, but the actual system output is the sum of the two parts. It might be thought that bounds on the RMS outputs of the two parts can be added together in an r.s.s. fashion; i.e., by taking the square root of the sum of their squares. Howver, this is not valid, since K could have sinusoidal components in phase with the output of C. Instead, a valid bound is achieved by adding the two bounds directly: if the input u is again $u = u_0 e^{j\omega t}$, where $u_0 \in C^n$, then

$$\|e\|_{\text{RMS}} \leq [\bar{\sigma}(C(j\omega)) + \bar{\sigma}(R(j\omega))] \, \|u_0\|_2$$

Proof:

$$\|e - Cu\|_\tau \leq \|Ru\|_\tau$$

$$\|e\|_\tau \leq \|e - Cu\|_\tau + \|Cu\|_\tau \leq \|Cu\|_\tau + \|Ru\|_\tau$$

$$\lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}} \|e\|_\tau \leq \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}} \|C(j\omega) u_0 e^{j\omega t}\|_\tau + \lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}} \|R(j\omega) u_0 e^{j\omega t}\|_\tau$$

$$\|e\|_{\text{RMS}} \leq [\bar{\sigma}(C(j\omega)) + \bar{\sigma}(R(j\omega))] \, \|u_0\|_2$$

Again, in the SISO case, if $u = k \sin \omega t$, where $k > 0$, then

$$\|e\|_{\text{RMS}} \leq (|C(j\omega)| + |R(j\omega)|) \frac{k}{\sqrt{2}} = (|C(j\omega)| + |R(j\omega)|) \|u\|_{\text{RMS}}$$

Examples of the ideas presented here are in the first example of Chapter 5.

4.3.4 RMS Output for White Noise Inputs

Some systems have performance specifications on the allowable RMS output error when white noise is applied to the input. This type of specification also translates easily into conic sector terms. If a system's input-output operator $\mathcal{K}$ is inside Cone$(0, R)$ and the input u is a stationary, ergodic, Gaussian white noise process satisfying $E[u(t+\tau)u^*(t)] = I\delta(\tau)$, $\forall\, t, \tau \in R$, then the RMS level of the output e. is

$$\|e\|_{\text{RMS}} \equiv \sqrt{E[\lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau e^*(\tau) e(\tau) d\tau]} = \sqrt{E[\lim_{\tau \to \infty} \frac{1}{\tau} \|e\|_\tau^2]} \leq \|R\|_2$$

where the 2-norm of a matrix-valued transfer function is defined by

$$\|R\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \|R(j\omega)\|_F^2 d\omega$$

This result follows from the definition of a conic sector. Assume that R is an $m \times n$ transfer function matrix. If $\mathcal{K}$ is inside Cone$(0, R)$, then for the u described above, $\|e\|_{\text{RMS}} \leq \|R\|_2$.

Proof: By the definition of conic sectors,

$$\|e\|_\tau^2 \leq \|Ru\|_\tau^2 \qquad \forall\, \tau > 0,\ \forall\, u \in L_{2e}$$

Therefore

$$\begin{aligned}
\|e\|_{\text{RMS}} &\equiv E[\lim_{\tau\to\infty}\frac{1}{\tau}\|e\|_\tau^2] \\
&\le E[\lim_{\tau\to\infty}\frac{1}{\tau}\|Ru\|_\tau^2] = \lim_{\tau\to\infty}E[\frac{1}{\tau}\int_0^\tau (Ru)^*(Ru)d\tau] \\
&= \lim_{\tau\to\infty}E\left[\frac{1}{\tau}\int_0^\tau \sum_{i=1}^m \left(\sum_{j=1}^n R_{ij}u_j\right)^* \left(\sum_{k=1}^n R_{ik}u_k\right) d\tau\right] \\
&= \sum_{i=1}^m\sum_{j=1}^n\sum_{k=1}^n \lim_{\tau\to\infty}E\left[\frac{1}{\tau}\int_0^\tau (R_{ij}u_j)^*(R_{ik}u_k)\,d\tau\right] \\
&= \sum_{i=1}^m\sum_{j=1}^n \lim_{\tau\to\infty}E\left[\frac{1}{\tau}\int_0^\tau |R_{ij}u_j|^2 d\tau\right] \\
&= \sum_{i=1}^m\sum_{j=1}^n \frac{1}{2\pi}\int_{-\infty}^{\infty}|R_{ij}(j\omega)|^2 d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\|R(j\omega)\|_F^2 d\omega = \|R\|_2^2
\end{aligned}$$

so the cone radius provides a direct bound on RMS output error in the white noise case as well as the sinusoidal input case. This time, the relevant quantity is the squared radius integrated over frequency, rather than the radius at a single frequency. Note that the condition $E[u(t+\tau)u^*(t)] = I\delta(\tau)$ implies that each element of the vector signal u is uncorrelated with the others, as well as being a white noise process of unit magnitude. If the actual input signal is frequency-dependent (colored) noise rather than white noise, or if the actual elements of u are correlated, it is always possible to construct an LTI operator which produces the actual signal when uncorrelated white noise is applied to its input. This operator can then be lumped with the system and the above test can still be applied.

If the best available conic sector description of the system has nonzero center, a bound analogous to the one in Section 4.3.3 can be used. Suppose $\mathcal{K}$ is inside $\text{Cone}(0, R)$, and the same white noise signal $u(t)$ satisfying $E[u(t+\tau)u(t)] = I\delta(\tau)$ is applied to its input. Then the RMS output level satisfies

$$\|e\|_{\text{RMS}} \le \|C\|_2 + \|R\|_2$$

This is true because

$$\begin{aligned}
\|e - Cu\|_\tau &\leq \|Ru\|_\tau \\
\|e\|_\tau &\leq \|e - Cu\|_\tau + \|Cu\|_\tau \leq \|Cu\|_\tau + \|Ru\|_\tau \\
E\left[\lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|e\|_\tau\right] &\leq E\left[\lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|Cu\|_\tau\right] + E\left[\lim_{\tau\to\infty} \frac{1}{\sqrt{\tau}}\|Ru\|_\tau\right] \\
\|e\|_{\text{RMS}} &\leq \|C\|_2 + \|R\|_2
\end{aligned}$$

Note that we have adopted very different definitions for the 2-norm depending on whether it is applied to a transfer function or a constant matrix. The 2-norm of a constant matrix is its maximum singular value; the squared 2-norm of a matrix transfer function is $1/2\pi$ times its squared Frobenius norm integrated over all frequencies.

If the radius $R(j\omega)$ does not roll off faster than $1/\sqrt{\omega}$ as $\omega \to \infty$, then the integral defining $\|R\|_2$ does not converge; it "blows up." In this case we can say nothing about the RMS response to white noise until a valid alternative radius if found. Recall that for sampled-data operators, the basic theorem of Section 3.1 produces cone radii that are periodic in ω, and do not roll off at all at high frequencies. Such a radius is unsuitable for this test. Fortunately, there can be much flexibility in choosing a valid radius R.

The best approach is to work from the specifications to determine what input and output frequency weights, and what cone center and radius C, R, are needed such that the frequency-weighted system will pass its specifications if it is inside Cone(C, R). Then apply the weights, subtract C, and weight the input further with R^{-1}. Next apply the tests of Sections 3.9.3 or 3.9.4 to see if the modified system is inside Cone$(0, I)$. The input weights should be chosen to convert the white noise $u(t)$ into the actual input seen by the system (colored noise, correlated inputs, magnitude other than 1). The output weights could be bandpass filters to account for specifications requiring, for instance, that the RMS output in a given

frequency band be less than a given amount.

4.3.5 Step Response

Systems are often given specifications on their step response, such as rise time, overshoot, decay time, or other time-domain quantitities. Often the specification is in the form of a template or envelope of allowable step responses; if the step response fits inside the envelope, it meets the specification. Unfortunately, conic sector methods do not lend themselves to time-domain envelope tests of this type. The reason is that the cone radius R puts bounds only on the squared output error integrated over all time, not on the value of the output at a single instant of time. That is, if a system K with input u and output y is inside $\mathrm{Cone}(C, R)$, we can say only that

$$\|y - Cu\|_\tau \equiv \sqrt{\int_0^\tau \|y - Cu\|_2^2 d\tau} \leq \|Ru\|_\tau \quad \forall\, u \in L_{2e},\ \forall \tau > 0.$$

It is possible, in principle, for $y - Cu$ to be a tall, narrow pulse similar to a δ-function and still pass this test. Since there is no limit on how tall the pulse can be, or on its position in time, we can make no categorical statements about the maximum output error for a given time t, unless we have additional information about the system.

This problem is not unique to conic sector methods, however, since it also occurs in conventional LTI analysis. If an LTI system's transfer function is uncertain and bounded by limits in the frequency domain, it is also impossible to guarantee that its step response is within any given envelope. The reason is the same: its step response could, in principle, contain a tall, narrow pulse. To make such guarantees, we would need more information such as the relative degree of the uncertain transfer function.

The cone center C can still be useful, however, for estimating the system's re-

sponse to a step or any other input. By replacing each sampler with $1/T_i$ and each LTI Δ-block with 0 in the manner already discussed, one can derive a transfer function C which gives a good approximation to the sampled-data system's response. Since a C obtained in this way contains both analog and digital elements, however, it will generally have no simple state-space representation.

One way to work around this is to sample $C(j\omega)$ on an evenly spaced grid of frequencies, then apply the Fast Fourier Transform (FFT) to find an approximation of the impulse response. The step response can then be formed by integrating the impulse response numerically.

Another approach is to use the complex cepstrum matching technique of Chapter 6 to find a good analog state-space approximation to C. Then conventional computer methods (numerical integration in state space) can be used to simulate the response of C to a step or to any other input waveform. An example of this appears in Chapter 5, where a systems's actual step response is compared to the approximate step response obtained from C.

Chapter Five — Examples

5.1 SISO Single Rate Example

Consider the feedback system of Figure 5.1-1, with elements given by

$$G(s) = \frac{150}{(s+1)(s+3)} \quad F(s) = \frac{2500}{s^2+70s+2500} \quad K(z) = \frac{.80498(z-.90993)^2}{(z-.98750)(z-.47744)}$$

with sampling period $T = .031416 = \pi/100$ sec and H_0^T a zero-order hold. $G(s)$ represents an analog plant, and $F(s)$ is a second order Butterworth prefilter with corner frequency at 50 rad/sec, or one-half the Nyquist frequency $\pi/T = 100$ rad/sec. This example has been used previously in [20] and [23]. We will use this example to demonstrate most of the analysis techniques presented in Chapters 3 and 4, including nominal stability, robust stability, nominal performance, robust performance, gain and phase margins, RMS gain for sinusoidal inputs, and step response approximation using the closed-loop cone center. In every test, the new methods prove to be extremely nonconservative.

Nominal stability. Rearrange the system as in Figure 5.1-2, with the sampler extracted outside the system and the 2×2 operator M representing the interconnection of the inputs and outputs of the sampler and the system itself. M is given by

$$M = \begin{bmatrix} -FGH_0^T K & FG \\ -GH_0^T K & G \end{bmatrix}$$

where $K(z)$ is always evaluated on the unit circle: $z = e^{j\omega T}$.

The next step, conceptually, is to sandwich the sampler between fictitious operators α and β and to cone-bound the result using Theorem 1. By lumping the cone

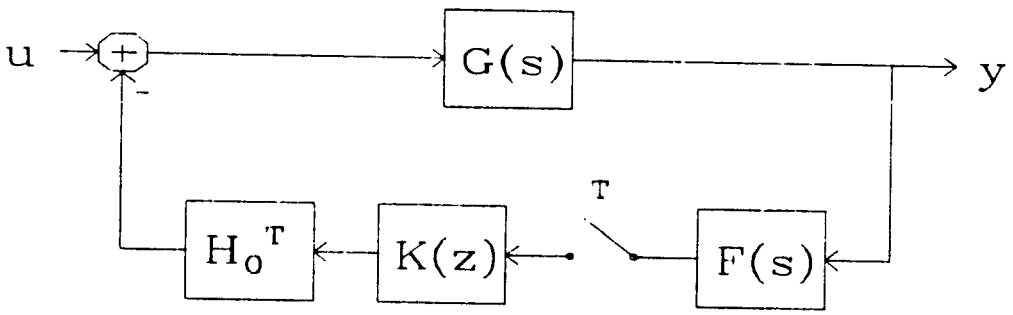

Fig. 5.1-1: A SISO single rate system.

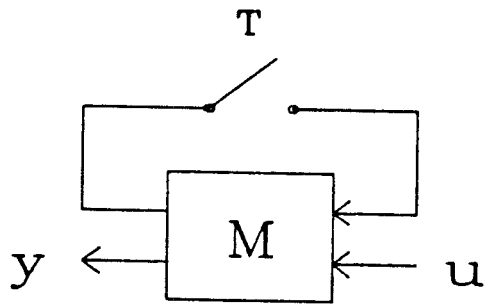

Fig. 5.1-2: The rearranged system with the sampler extracted from the 2×2 operator M.

center with the system M to produce $\tilde{M}$, and breaking up the remaining zero-center operator into $R^{1/2}\hat{\Delta}R^{1/2}$, where $\hat{\Delta}$ is a PLTV operator inside Cone$(0, I)$, we get the transformed system of Figure 5.1-3. $\tilde{M}$ is given by

$$\tilde{M} = \begin{bmatrix} -FGH_0^TK(1+\frac{1}{T}FGH_0^TK)^{-1} & FG(1+\frac{1}{T}H_0^TKFG)^{-1} \\ -GH_0^TK(1+\frac{1}{T}FGH_0^TK)^{-1} & G(1+\frac{1}{T}H_0^TKFG)^{-1} \end{bmatrix} = \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix}$$

Actually, we only need $\tilde{M}_{11}$ to perform the nominal stability test, but we will use all of $\tilde{M}$ to make the nominal performance test next. Following Section 3.4, we choose the sandwiching operators α and β to give an optimally small value for the stability criterion $\|R\tilde{M}_{11}/\alpha\beta\|_\infty$:

$$\alpha(\omega) = \beta(\omega) = |\tilde{M}_{11}(j\omega)|^{1/2}$$

Next compute $R = \text{Rad}(\alpha, \beta)$; now since $|R\tilde{M}_{11}/\alpha\beta| = |R|$, a plot of the cone radius tells us if the stability test is satisfied. Figure 5.1-4 shows R, and it can be seen that $\|R\|_\infty$ is much less than 1: actually $\|R\|_\infty = .0052$. So this part of the nominal stability test is passed by a wide margin.

The other part of the nominal stability test is simply to verify that $\tilde{M}_{11}$ is stable, or equivalently that the system is stable when the sampler is replaced by $1/T$. Figure 5.1-5 shows a log Nyquist plot of the cone center loop gain $\frac{1}{T}FGH_0^TK$; it can be seen that the Nyquist locus makes no encirclements of the critical point -1, so $\tilde{M}_{11}$ is stable and the system passes both parts of the nominal stability test.

The log Nyquist plot in Figure 5.1-5 is of the same type introduced in Section 3.7.2; the distance from the origin is $1 + \log_{10}(|\frac{1}{T}FGH_0^TK|)$ at each frequency for points larger in magnitude than 0.1. Points smaller than 0.1 in magnitude are mapped to the origin. Phase is represented in the same way as in a conventional Nyquist plot.

Figure 5.1-5 actually shows two log Nyquist plots, but they are so close together that they are indistinguishable. The other plot is of the discretized loop gain

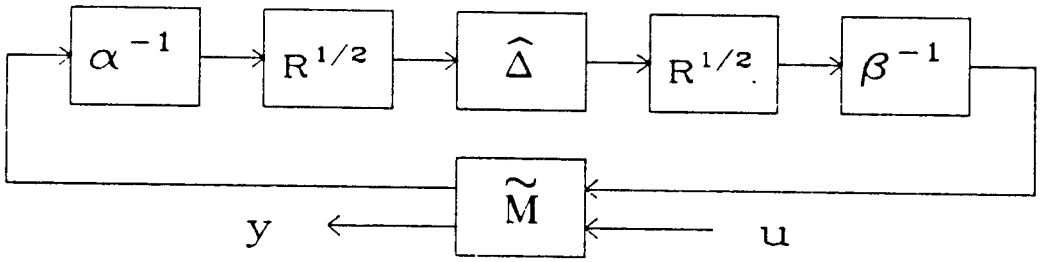

Fig. 5.1-3: The system after sandwiching and cone bounding the sampler, then lumping the cone center with M to get $\tilde{M}$. $\hat{\Delta}$ is inside Cone$(0, I)$.

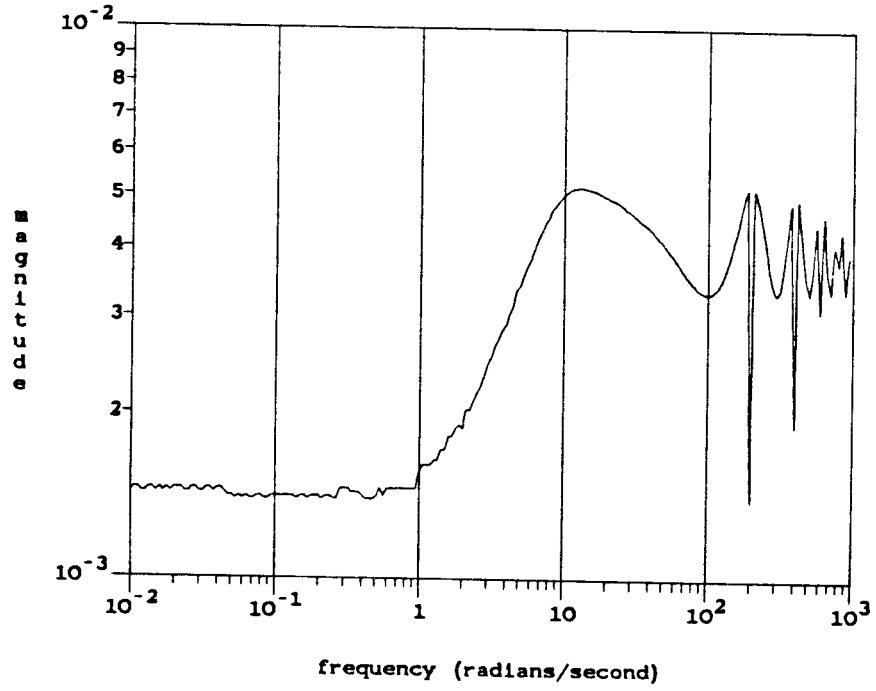

Fig. 5.1-4: Bode plot of cone radius $R = \text{Rad}(\alpha, \beta)$. Since $\|R\|_\infty = .0052 < 1$, the system passes part of the nominal stability test.

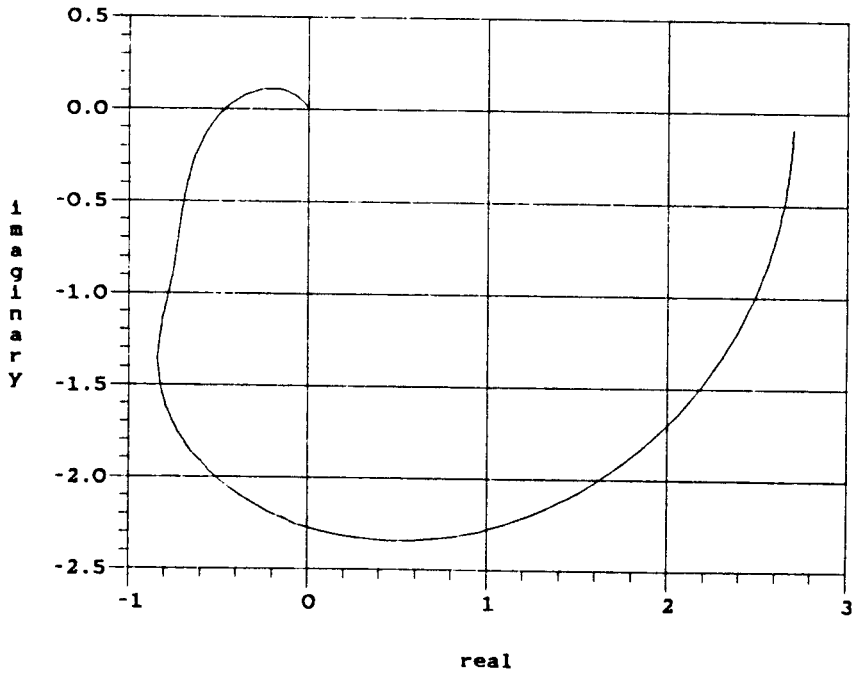

Fig. 5.1-5: Log Nyquist plot of the cone center loop gain $\frac{1}{T}FGH_0^TK$. Since it does not encircle -1, $\tilde{M}$ is stable and the system passes the second part of the nominal stability test. Also plotted is the discretized loop gain $K(FGH_0^T)^T$ from 0 to π/T. The plots are indistinguishable.

$K(z)(FGH_0^T)^T(z)$ from 0 to π/T in frequency. Since this Nyquist locus also makes no encirclements of -1, the conventional method also indicates nominal stability for the system. The two plots are nearly identical because the steep high frequency rolloff of FGH_0^T keeps aliasing to a minimum when $(FGH_0^T)^T$ is computed. Actually the plots are quite different at high frequencies, since the discretized loop gain is periodic in ω, and the cone center loop gain decays to 0 as $\omega \to \infty$.

Nominal performance. In Section 3.9.3, a sampled-data system is said to meet its nominal performance requirements if its closed loop input-output response is inside Cone$(0, I)$. By applying suitable frequency weights to the system input and adding an LTI operator C in parallel to the system, this problem can be made equivalent to testing whether the closed-loop system is inside a given Cone(C, R_2). Since we have no such performance specification for this example, we can run the problem in reverse and find a Cone(C, R_2) which contains the closed-loop response and which has the smallest possible radius R_2.

We can, in principle, choose any operator C for our cone center, but a good choice is $C = \tilde{M}_{22} = G(1 + \frac{1}{T}H_0^T K F G)^{-1}$. This is the closed-loop response of the LTI system produced when the sampler is replaced by $1/T$. Now construct the system of Figure 5.1-6. $\tilde{M}$,α, β, $R^{1/2}$, and $\hat{\Delta}$ are the same operators as in Figure 5.1-3. In addition, operators D and D^{-1} have been sandwiched around the PLTV operator $\hat{\Delta}$, C has been subtracted from the response from input 2 to output 2 of $\tilde{M}$, and a new weighting operator R_2^{-1} has been placed at input 2 of the new system $\hat{M}$. $\hat{M}$ is the 2×2 LTI operator inside the dashed box; it is given by

$$\hat{M} = \begin{bmatrix} \frac{R\tilde{M}_{11}}{\alpha\beta} & \frac{DR^{1/2}\tilde{M}_{12}}{\alpha R_2} \\ \frac{R^{1/2}\tilde{M}_{21}}{\beta D} & \frac{\tilde{M}_{22} - C}{R_2} \end{bmatrix} = \begin{bmatrix} \hat{M}_{11} & \hat{M}_{12} \\ \hat{M}_{21} & \hat{M}_{22} \end{bmatrix}$$

$\hat{M}_{22}$ goes to zero when we set $C = \tilde{M}_{22}$; this is an indication that this is a good choice for C. The purpose of C and R_2 is to convert a test for Cone(C, R_2) into a test for Cone$(0, I)$, since it is an elementary conic sector result that for any operator

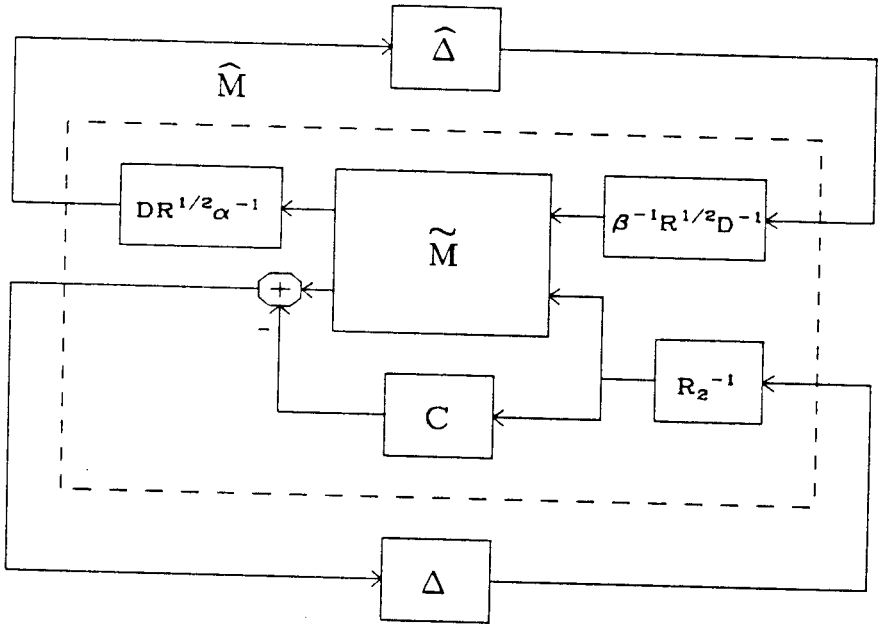

Fig. 5.1-6: $\hat{M}$ for the nominal performance test. If this system is stable for all Δ in Cone$(0, I)$, then the closed-loop response is inside Cone(C, R_2).

K,

$$K \text{ inside Cone}(C, R_2) \Leftrightarrow (K - C)R_2^{-1} \text{ inside Cone}(0, I).$$

Now by applying the nominal performance test of Section 3.9.3 to $\hat{M}$ and by decreasing R_2 at each frequency until the system fails the test, we may find the smallest radius R_2 such that the nominal closed loop system in inside Cone(C, R_2).

A word must be said about the LTI operator D. Its inclusion in the system is valid only if $D\hat{\Delta}D^{-1} = \hat{\Delta}$. Since $\hat{\Delta}$ is PLTV with period T, this will be true only when D is periodic in ω with period $2\pi/T$. If $\mathcal{D}$ is the set of such operators, we want to compute the smallest R_2 such that

$$\left\| \inf_{D \in \mathcal{D}} \bar{\sigma}(\hat{M}(j\omega)) \right\|_\infty \leq 1$$

With D constrained to be periodic, it is not obvious how to make this calculation. We can compromise by selecting a D which is *constant* over frequency. A good way to choose this is to make a first computer run with no constraints on D by computing, at each frequency, the smallest R_2 such that $\inf_{D \in \mathcal{D}} \bar{\sigma}(\hat{M}(j\omega)) = 1$, where $\mathcal{D}$ now includes non-periodic transfer functions. Then find the frequency at which R_2 is largest, and set D to be constant at its value at that frequency. Then make a second computer run with D constant, and compute the smallest R_2 at each frequency such that $\bar{\sigma}(M(j\omega)) = 1$.

Figure 5.1-7 shows the result of such a computation: a conic sector Cone(C, R_2) containing the nominal closed loop response from u to y. The magnitude of both R_2 and $C = \hat{M}_{22} = G(1 + \frac{1}{T}H_0^T KFG)^{-1}$ are shown. Note that the cone radius is much smaller than the cone center from DC to about 1000 rad/sec; this indicates that the closed loop system acts much like an LTI system, with little distortion due to aliasing or "stairstepping."

Robust stability. To test robust stability, we must first define the structure of the plant uncertainty. Consider Figure 5.1-8, in which multiplicative uncertainty

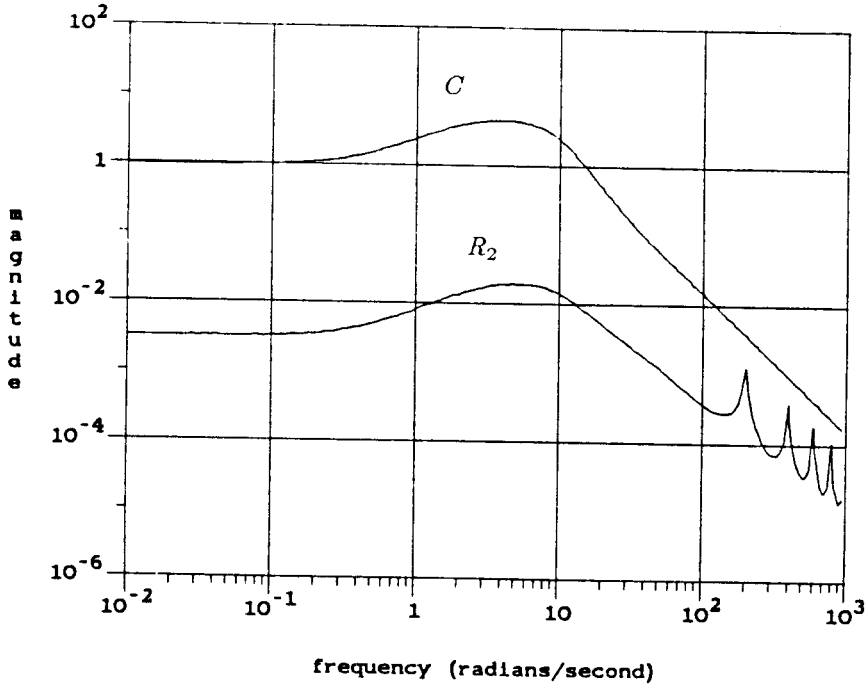

Fig. 5.1-7: Bode plots of C and R_2 for the nominal performance test. The closed-loop response is inside Cone(C, R). Since $|R| \ll |C|$ at all frequencies, the closed-loop system acts approximately like the LTI operator $C = G(1 + \frac{1}{T} H_0^T K F G)^{-1}$.

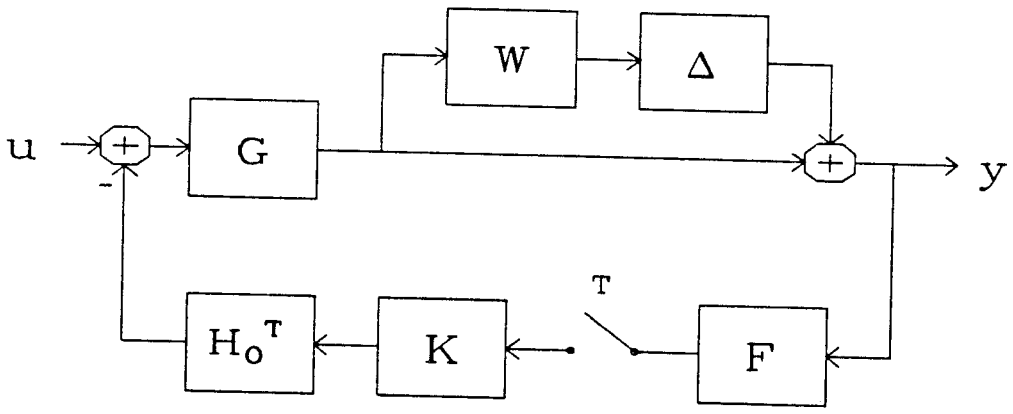

Fig. 5.1-8: The system with multiplicative uncertainty $1 + \Delta W$ at the output of G.

$1 + W\Delta$ has been inserted at the output of G. If we had a specification for the allowable amount of perturbation $|W\Delta|$ at each frequency, we could set $|W|$ equal to the maximum allowed perturbation and then make the robust stability test with $\|\Delta\|_\infty \leq 1$. Instead, we will work in reverse to find the largest allowable $|W\Delta|$ at each frequency which maintains stability.

Redraw the system as in Figure 5.1-9, with the sampler extracted above and the Δ-block below the interconnection matrix M. M is given by

$$M = \begin{bmatrix} -FGH_0^T K & F \\ -GH_0^T K & 0 \end{bmatrix}$$

Now by sandwiching the sampler between α and β as before, applying Theorem 1, lumping the cone center with M, and factoring the remaining PLTV operator into $R^{1/2}\hat{\Delta}R^{1/2}$ as before, where $\hat{\Delta}$ is inside Cone$(0, I)$, we get the system of Figure 5.1-10. $\tilde{M}$ is the new interconnection matrix with the cone center lumped with M, and $\hat{M}$ is the 2×2 LTI operator in the dashed box. $\tilde{M}$ is given by

$$\tilde{M} = \begin{bmatrix} -FGH_0^T D(1+L_c)^{-1} & F(1+L_c)^{-1} \\ -GH_0^T D(1+L_c)^{-1} & -\frac{1}{T}GH_0^T DF(1+L_c)^{-1} \end{bmatrix} = \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix}$$

where $L_c = \frac{1}{T}FGH_0^T K$ is the loop gain when the sampler is replaced by $1/T$. $\hat{M}$ is given by

$$\hat{M} = \begin{bmatrix} \frac{R\tilde{M}_{11}}{\alpha\beta} & \frac{R^{1/2}\tilde{M}_{12}}{\alpha D} \\ \frac{R^{1/2}\tilde{M}_{21}DW}{\beta} & \tilde{M}_{22}W \end{bmatrix}$$

Now by the robust stability method of Section 3.9.4, this system is stable for all stable LTI Δ satisfying $\|\Delta\|_\infty < 1$ if $\|\hat{M}\|_\infty \leq 1$. Since D and D^{-1} are sandwiched around an LTI operator this time, not a PLTV operator, there is no constraint that D must be periodic in ω. If $\mathcal{D}$ is the set of all LTI operators, then, we can compute $\inf_{D\in\mathcal{D}} \bar{\sigma}(\hat{M})$ at each frequency. By finding the largest W at each frequency such that $\inf_{D\in\mathcal{D}} \bar{\sigma}(\hat{M}) = 1$, we can find a frequency dependent robustness margin. Figure 5.1-11 shows a magnitude Bode plot of the margin obtained in this way. The system is guaranteed stable for any multiplicative perturbation lying below that curve.

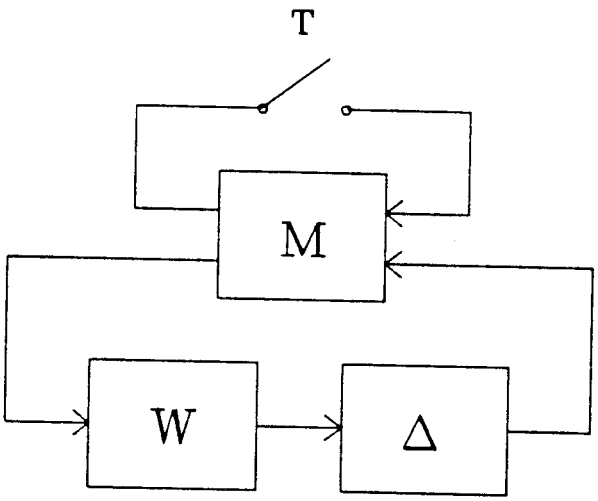

Fig. 5.1-9: The restructured system for the robust stability test.

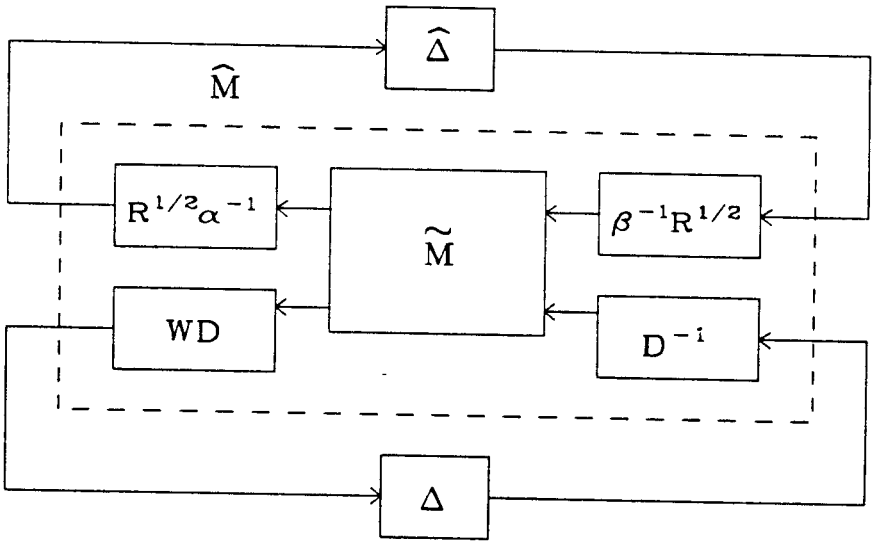

Fig. 5.1-10: Construction of $\hat{M}$ for the robust stability test. The $\hat{\Delta}$ and Δ blocks are both inside Cone$(0, I)$; Δ is LTI and $\hat{\Delta}$ is PLTV with period T.

For comparison purposes, Figure 5.1-11 also shows $|1 + L_d^{-1}(z)|$ over frequency, where $L_d(z) = K(z)(FGH_0^T)^T(z)$ is the discretized loop gain. Conventional digital analysis techniques tell us that the system will be stable for any *constant* (not frequency dependent) multiplicative perturbation Δ which satisfies $|\Delta| < \inf_{\omega \in R} |1 + L_d^{-1}(e^{j\omega T})|$. The two curves are nearly equal at their minima: the minimum of $|1 + L_d^{-1}(z)|$ is .7445 at $\omega = 6.31$ rad/sec, and the value of the conic sector robustness margin at the same frequency is .7417, only 0.4% smaller. The conic sector robust stability test, then, is extremely nonconservative for this example. Furthermore, it gives us much more usable information than the conventional test, since it is a frequency dependent quantity. For instance, at $\omega = 100$, it shows that a multiplicative perturbation with magnitude of nearly 200 is acceptable, while the digital test implies that multiplicative perturbations should be less than .7445 in magnitude at all frequencies.

This illustrates an important motivation for using conic sector methods: they preserve the analog nature of a sampled-data feedback system. Since models for physical systems typically have small uncertainty at low frequencies and large uncertainty at high frequencies, it is important to be able to derive robustness margins which reflect this fact.

Robust performance. For the robust performance test, we can use the same type of plant uncertainty as for the robust stability test: the system of Figure 5.1-8, with a multiplicative perturbation $1 + W\Delta$ at the output of G. This time, we want to find a conic sector, with the smallest possible radius, which contains the closed loop response from u to y for any perturbation $W\Delta$ in a given set. This time we will work in the forward direction by specifying the set of perturbations first.

For this example, let $W = 0.3$ and $|\Delta(j\omega)| < 1$ at all frequencies; this is 30% plant uncertainty. A more realistic choice for W might increase in magnitude at high frequencies, but this choice is sufficient to illustrate the method.

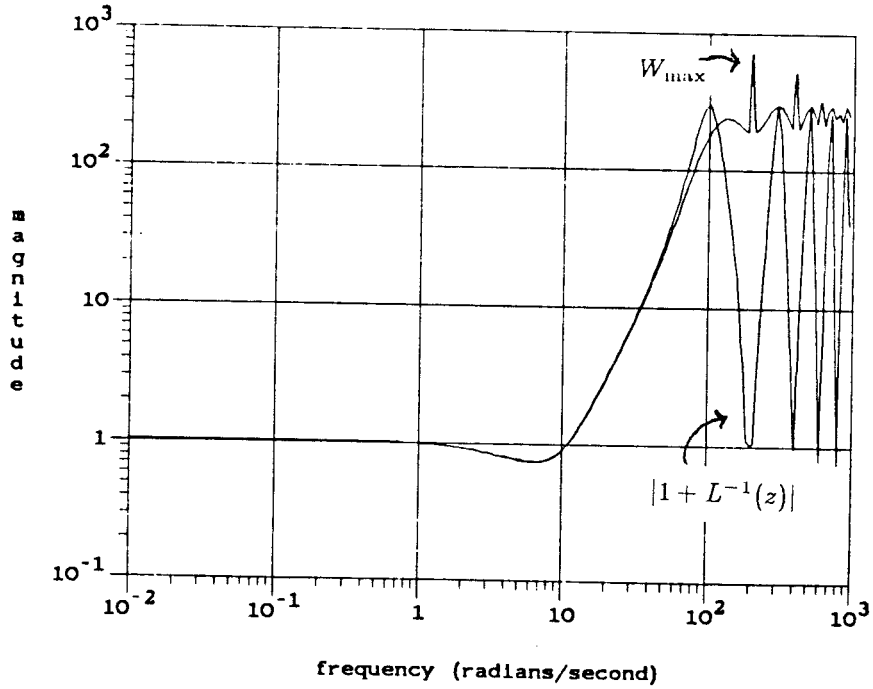

Fig. 5.1-11: Bode plots of $|1 + L^{-1}(z)|$ and the largest W which passes the robust stability test. At their minima, W is only 0.4% smaller than $|1 + L^{-1}(z)|$, indicating nonconservativeness.

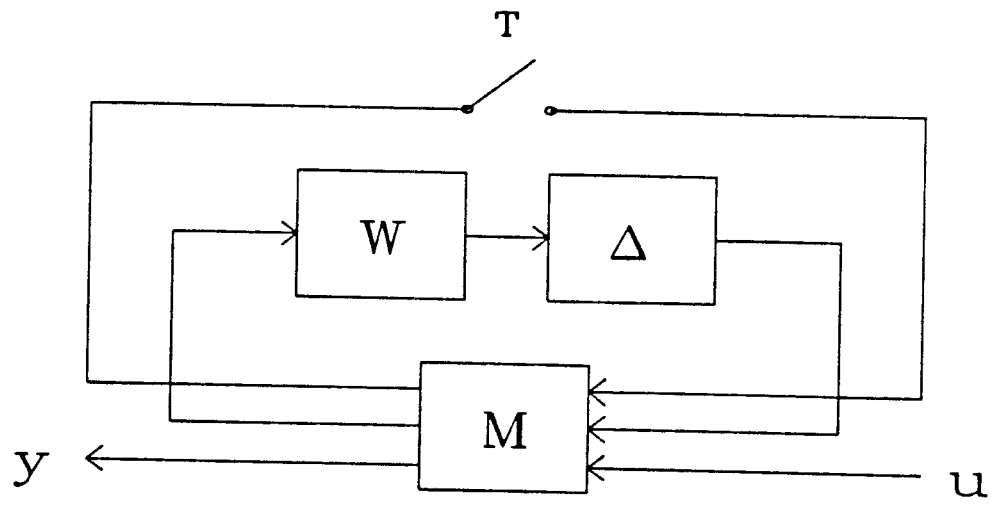

Fig. 5.1-12: The restructured system for the robust performance test.

Redraw the system as in Figure 5.1-12 with the sampler and $W\Delta$ both extracted above the system. The 3×3 interconnection matrix is given by

$$M = \begin{bmatrix} -FGH_0^T K & F & FG \\ -GH_0^T K & 0 & G \\ -GH_0^T K & 1 & G \end{bmatrix}$$

Now by folding the sampler's cone center $1/T$ into the system in the usual manner, we get the closed loop interconnection matrix $\tilde{M}$:

$$\tilde{M} = \begin{bmatrix} -FGH_0^T(1+L_c)^{-1} & F(1+L_c)^{-1} & FG(1+L_c)^{-1} \\ -GH_0^T K(1+L_c)^{-1} & -\frac{1}{T}GH_0^T KF(1+L_c)^{-1} & G(1+L_c)^{-1} \\ -GH_0^T K(1+L_c)^{-1} & (1+L_c)^{-1} & G(1+L_c)^{-1} \end{bmatrix}$$

where $L_c = \frac{1}{T}FGH_0^T K$ is the loop gain when the sampler is replaced by $1/T$.

Figure 5.1-13 shows the system with all three input-output loops closed; through $\hat{\Delta}$, the PLTV operator inside Cone$(0, I)$; through Δ, the uncertain LTI operator satisfying $\|\Delta\|_\infty < 1$; and through a new fictitious Δ_2 operator inside Cone$(0, I)$. As explained in Section 3.9.4, we can make no assumptions such as linearity for Δ_2; we assume only that it is inside Cone$(0, I)$.

If the system of Figure 5.1-13 is stable for all Δ_2 in Cone$(0, I)$, then the response from input 3 to output 3 of the operator $\hat{M}$ in the dashed box is also inside Cone$(0, I)$; this was shown in Section 3.9.3. This, in turn, is equivalent to the original system's closed-loop response from u to y being inside Cone(C, R_2); this fact was already used to test nominal performance above. We will set $C = \tilde{M}_{33} = G(1 + \frac{1}{T}H_0^T KFG)^{-1}$ and then find the smallest R_2 at each frequency which passes this test. The test requires that $\bar{\sigma}(\hat{M}(j\omega)) < 1$ at all frequencies.

We can take advantage of our knowledge of $\hat{\Delta}$ and Δ to include the scaling factors D_1 and D_2, which can be varied to reduce $\bar{\sigma}(\hat{M})$. Since Δ is LTI, any choice for D_2 satisfies $D_2^{-1}\Delta D_2 = \Delta$; therefore D_2 is a free parameter at every frequency. $\hat{\Delta}$, however, is PLTV; therefore $D_1^{-1}\hat{\Delta}D_1 = \hat{\Delta}$ only if $D_1(j\omega)$ is periodic in ω with period $2\pi/T$. Rather than work with this messy constraint, we will instead choose D_1 to be constant over frequency.

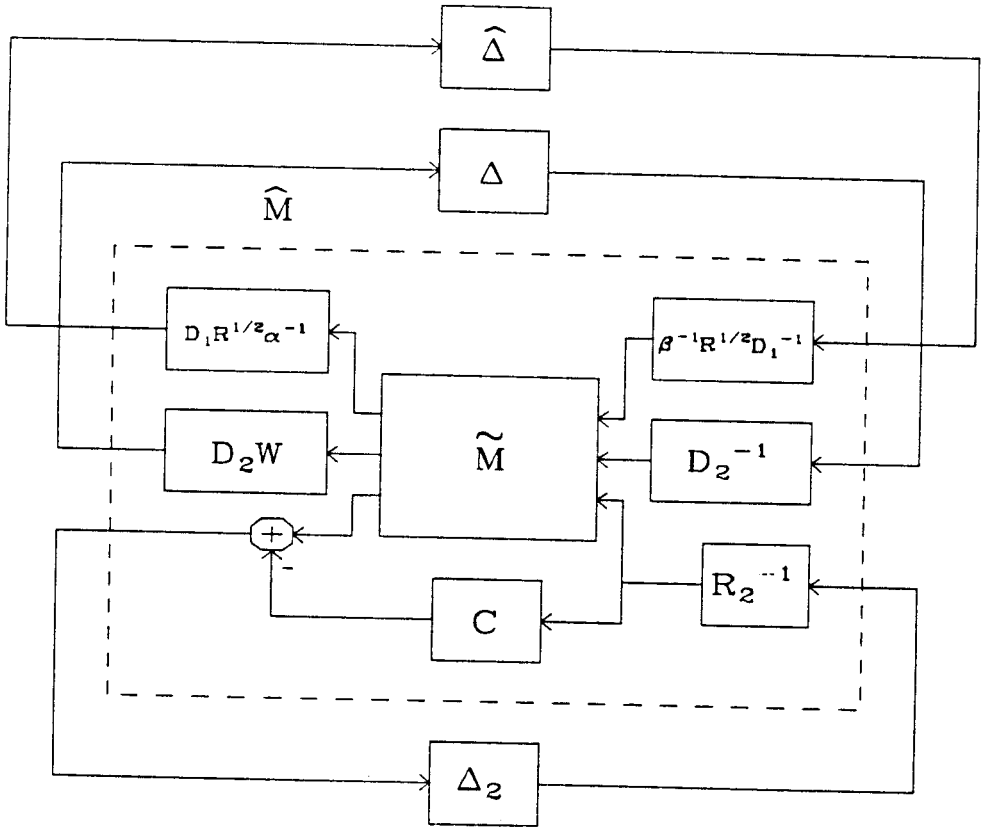

Fig. 5.1-13: Construction of $\hat{M}$ for the robust performance test. Δ, $\hat{\Delta}$, and Δ_2 are all inside Cone$(0, I)$. Δ is LTI and $\hat{\Delta}$ is PLTV with period T. If this system is stable for all Δ_2 inside Cone$(0, I)$, the closed-loop response is inside Cone(C, R_2).

The full expression for $\hat{M}$ is

$$\hat{M} = \begin{bmatrix} \frac{R_1\tilde{M}_{11}}{\alpha\beta} & \frac{D_1 R_1^{1/2}\tilde{M}_{12}}{D_2\alpha} & \frac{D_1 R_1^{1/2}\tilde{M}_{13}}{R_2\alpha} \\ \frac{D_2 W R_1^{1/2}\tilde{M}_{21}}{D_1\beta} & W\tilde{M}_{22} & \frac{D_2 W\tilde{M}_{23}}{R_2} \\ \frac{R_1^{1/2}\tilde{M}_{31}}{D_1\beta} & \frac{\tilde{M}_{32}}{D_2} & \frac{\tilde{M}_{33}-C}{R_2} \end{bmatrix}$$

We can solve the problem in two steps, similarly to the way the nominal performance test was made. Make a first computer run with both D_1 and D_2 free parameters at each frequency, and find the smallest R_2 at each frequency such that $\bar{\sigma}(\hat{M}) = 1$. Then set D_1 to its value at the frequency where the resulting R_2 is largest. Make another computer run with D_1 constant at this value, and find the smallest R_2 so that $\bar{\sigma}(\hat{M}) = 1$ with D_2 still free. This value of $R_2(j\omega)$ is a valid cone radius for the closed loop system for *all* stable perturbations $\|\Delta\|_\infty < 1$.

Figure 5.1-14 shows a plot of C and R_2 computed in this way. C is, of course, the same cone center used for the nominal performance test. R_2 is now significantly larger at all frequencies than it was in the nominal case, reflecting the added plant uncertainty. R_2 is still smaller than C for $\omega < 200$, indicating that the system will still act much like an LTI operator, with little distortion from aliasing or stairstepping.

Gain and phase margins. We may compute the exact gain and phase margins for this system by conventional LTI methods, since there is only a single sampler. Since Figure 5.1-5 shows the discretized loop gain $L_d(z) = K(z)(FGH_0^T)^T(z)$ (as well as the cone center loop gain $L_c = \frac{1}{T}FGH_0^T K$), the gain and phase margins may be read off this plot. The loop gain crosses the real axis at the point $-.2931$ at 17.8 rad/sec, so the gain can be increased by a factor of $1/.2931 = 3.412$ before the locus hits the critical point -1; therefore the gain margin is

$$\text{GM} = [0, 3.41] = [-\infty \text{ dB}, +10.7 \text{ dB}]$$

The loop gain locus crosses the unit circle at 6.69 rad/sec with a phase shift of

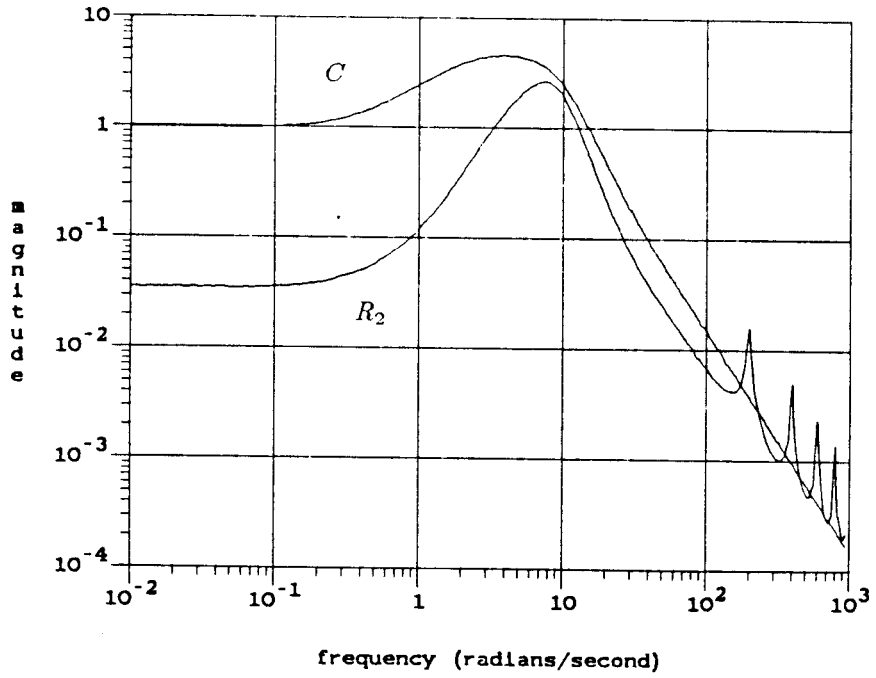

Fig. 5.1-14: Bode plots of C and R_2 for the robust performance test. The perturbed closed-loop system is inside Cone(C, R_2).

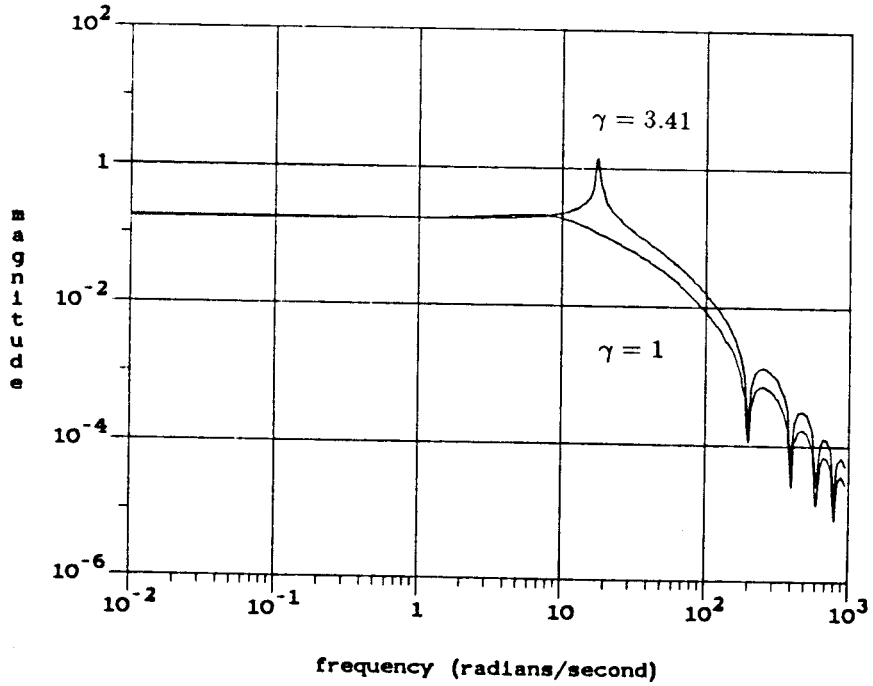

Fig. 5.1-15: Behavior of the optimal sandwiching operator α when G is scaled by γ. When $\gamma \approx 3.41$, α develops a large peak, indicating a gain margin of about 3.41.

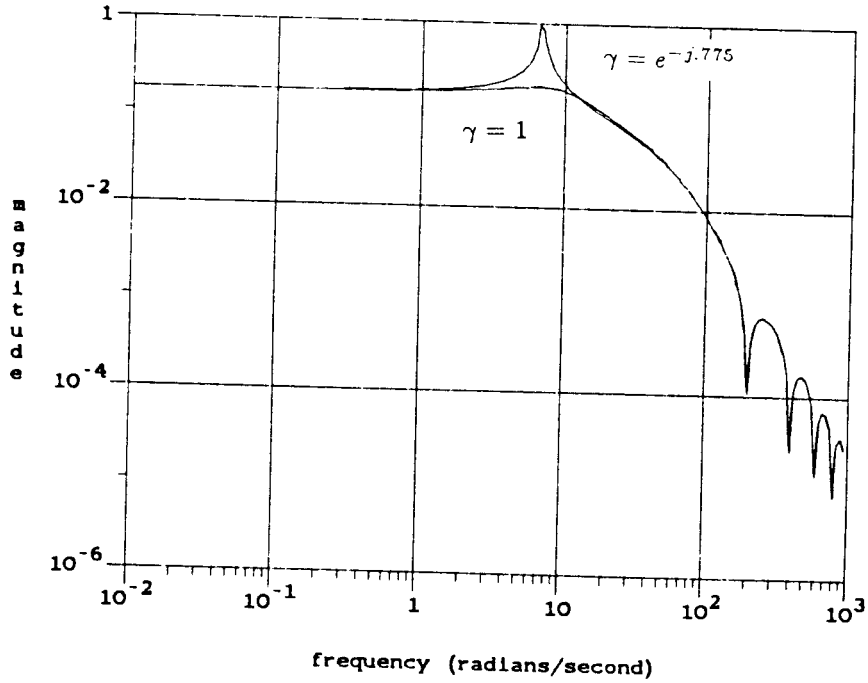

Fig. 5.1-16: Behavior of α when G is scaled by γ. When $\gamma \approx e^{-j.775}$, α develops a large peak, indicating a phase margin of about 44°.

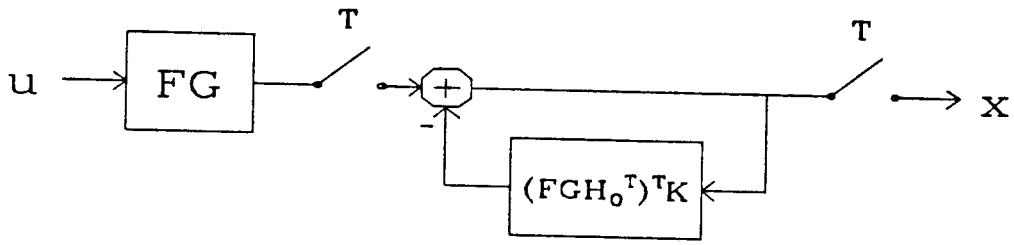

Fig. 5.1-17: Restructured system for computing exact response to sine inputs.

-135.6^o; so the phase can be rotated by 44.4^o before crossing the critical point. The phase margin is therefore

$$\text{PM} \;=\; [-44.4^o, +44.4^o]$$

We can find guaranteed gain and phase margins (sufficient conditions only) by conic sector methods. Recall that for the nominal stability test, the optimal sandwiching operators α and β are given by

$$\alpha(\omega) \;=\; \beta(\omega) \;=\; \left|\tilde{M}_{11}(j\omega)\right|^{1/2} \;=\; \left|FGH_0^T K(1+\frac{1}{T}FGH_0^T K)^{-1}\right|^{1/2}$$

If a fictitious gain block γ is placed in series with G, the new optimal α and β are

$$\alpha(\omega) \;=\; \beta(\omega) \;=\; \left|\gamma FGH_0^T K(1+\frac{1}{T}\gamma FGH_0^T K)^{-1}\right|^{1/2}$$

By varying γ in the neighborhoods of $\gamma = 3.412$ and $\gamma = e^{-j2\pi(44.4/360)}$ until the system fails the nominal stability test, we can generate lower bounds for the gain and phase margins. Recall that the test is

$$\left|\frac{R\tilde{M}_{11}}{\alpha\beta}\right| \;=\; |R| \;\leq\; 1 \qquad \forall\, \omega \in R$$

where $R = \text{Rad}(\alpha, \beta)$. Figures 5.1-15 and 5.1-16 show what happens when γ takes on values in these critical ranges: α and β acquire strong peaks at the frequency where the loop gain approaches the critical point. The magnitudes of the peaks rise to infinity as γ approaches the critical values 3.412 and $e^{-j.775}$; the radius R also acquires peaks at the same frequencies.

However, the value of γ gets so close to the critical values before $\|R\|_\infty > 1$ that it is beyond the accuracy of these computer programs to measure how close. The reason is that all curves are evaluated only at discrete points, not on a continuum, so the points never coincide perfectly with the critical point -1. Therefore we can only say that the gain and phase margins computed by conic sector methods are

negligibly smaller than the actual margins found by LTI methods. Conic sector methods are extremely nonconservative for this example.

RMS output for sine inputs. In Section 4.3.3 it was shown that if a sine wave u is input to a SISO operator inside Cone(C, R), then the RMS level of the output y satisfies

$$\frac{\|y\|_{\text{RMS}}}{\|u\|_{\text{RMS}}} \leq |C(j\omega)| + |R(j\omega)|$$

where ω is the frequency of u. We can compute an exact expression for the RMS output of this system when u is a sine wave, and compare it to $(|C| + |R|)\|u\|_{\text{RMS}}$ to test the conservativeness of this bound.

Redraw the system as in Figure 5.1-17, where the output of the sampler is extracted as the system output, rather than y. The output of this modified system, and therefore of the sampler in the original system, is

$$x = (1 + L_d(z))^{-1}(FGu)^T \quad \text{where } L_d(z) = K(z)(FGH_0^T)^T(z)$$

If $u = e^{j\omega t}$, we may use the identity $(u)^T = \frac{1}{T}\sum_k e^{j(\omega + k\omega_s)t}$, where $\omega_s = 2\pi/T$, to write

$$x = (1 + L_d)^{-1}FG\frac{1}{T}\sum_k e^{j(\omega+k\omega_s)t}$$

The output y of the original system is

$$\begin{aligned} y &= Gu - GH_0^T Kx \\ &= Ge^{j\omega t} - FG(1 + L_d)^{-1}\frac{1}{T}\sum_k (GH_0^T K)_k e^{j(\omega+k\omega_s)t} \\ &= \left[G - \frac{1}{T}FG(1 + L_d)^{-1}GH_0^T K\right]e^{j\omega t} - \frac{1}{T}FG(1 + L_d)^{-1}\sum_{k\neq 0}(GH_0^T K)_k e^{j(\omega+k\omega_s)t} \end{aligned}$$

and the squared RMS level is found by adding the squared magnitude of each component of y:

$$\|y\|_{\text{RMS}}^2 = |G|^2\left|1 - \frac{1}{T}FGH_0^T K(1 + L_d)^{-1}\right|^2 + \left|\frac{1}{T}FG(1 + L_d)^{-1}\right|^2 \sum_{k\neq 0}\left|(GH_0^T K)_k\right|^2$$

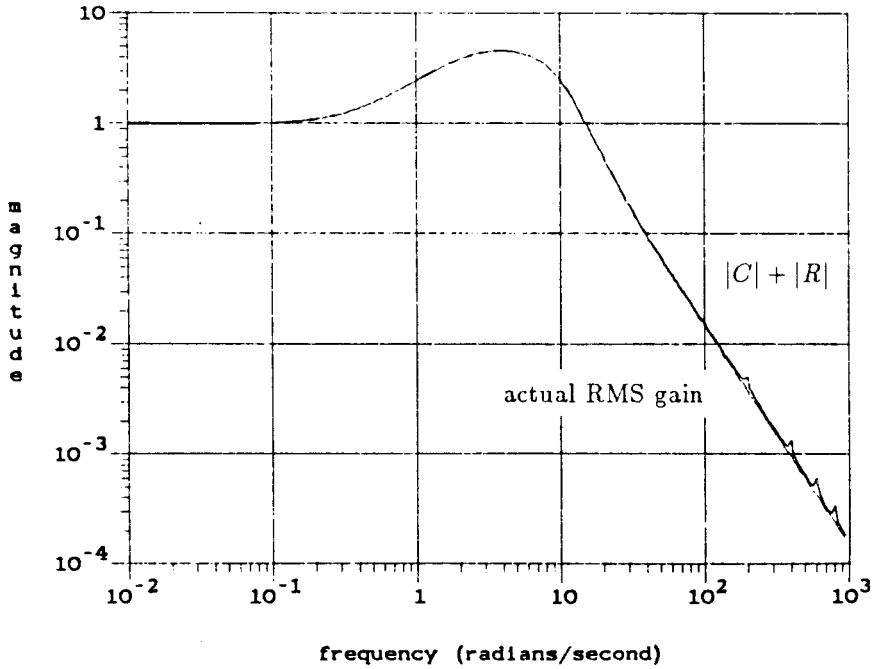

Fig. 5.1-18: Bode plots of actual RMS gain for sine inputs and the upper bound $|C| + |R|$. The conic sector bound is very tight, only 0.1% larger at low frequencies.

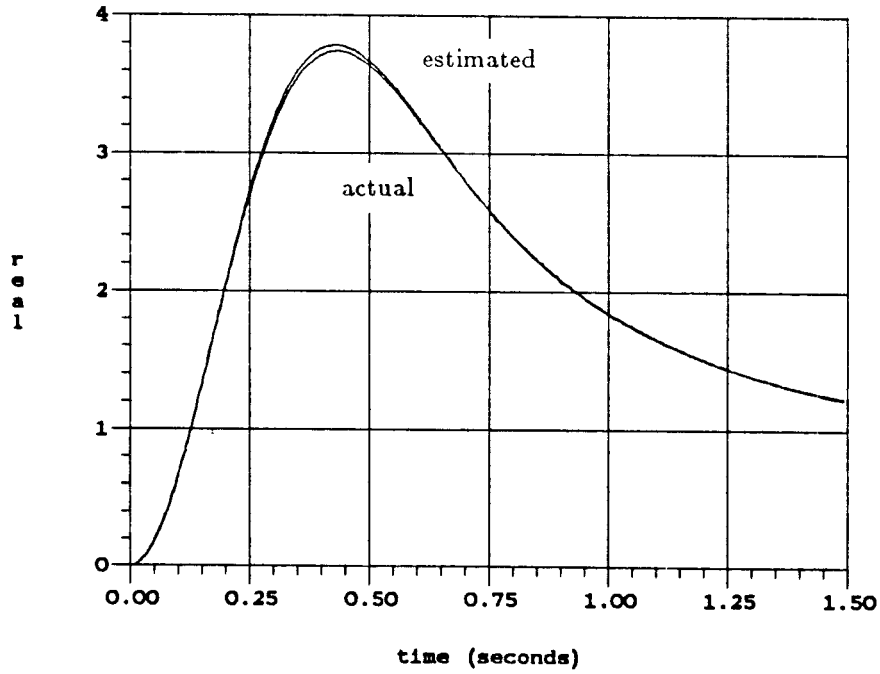

Fig. 5.1-19: Actual step response and estimated step response based on cone center C. A state-space representation for C was found using the complex cepstrum method of Chapter 6.

where $(GH_0^T K)_k \equiv (GH_0^T K)(j(\omega + k\omega_s))$. Since $\|u\|_{\text{RMS}}^2 = 1$, this expression is the actual squared RMS gain of the system. Figure 5.1-18 is a plot of this actual RMS gain and the upper bound $|C| + |R|$, where C and R are the cone center and radius derived earlier in the nominal performance test. The upper bound is so tight that the curves are nearly indistinguishable, except for small peaks at multiples of 200 rad/sec where R has peaks. A closer check of the computed values shows that $|C| + |R|$ is greater than the actual RMS gain at every frequency (as it should be!), but only by a very small difference of 0.1% at low frequencies. Conic sector methods are very nonconservative in this example.

Step response. Section 4.3.5 explained why conic sector methods cannot provide error bounds in the time domain, such as an envelope containing the step response. Nevertheless, the closed-loop cone center can be useful in computing approximate simulations of the closed-loop system, as long as the radius is small by comparison.

One difficulty appears because the cone center typically is a mixture of analog and digital transfer functions; therefore its state space is infinite-dimensional. The cone center for this example is

$$C(j\omega) \;=\; G(j\omega)\left(1 + \frac{1}{T} H_0^T(j\omega) K(e^{j\omega T}) F(j\omega) G(j\omega)\right)^{-1}$$

The discrete-time operator K causes the problem in this case. The problem may be circumvented by using the complex cepstrum matching technique of Chapter 6 to approximate $C(j\omega)$ with a finite-degree analog transfer function. A state-space quadruple can be constructed for the resulting approximation of C. The approximation's step response can then be computed by numerical integration in state space or by direct Laplace transform inversion of the transfer function.

This same $C(j\omega)$ is used as an example of the complex cepstrum matching method in Chapter 6. The result is a fifth order transfer function which approx-

imates $C(j\omega)$ so well that the difference is not visible on a Bode plot. The step response of this approximation is shown in Figure 5.1-19. Also shown is the actual system step response at $T/10$ time intervals, which was computed by the Kranc operator method of Section 2.8.

The two plots are very close, differing significantly only by a small error at the peak of the overshoot. This shows that the cone center provides a very good model of the closed-loop system in this example.

Note that even though the output in Figure 5.1-19 was computed with 10 points for every sampling period T, none of the stairstepping typical of many sampled data systems is apparent. This is due to the fact the G imposes a second order high frequency rolloff between the output of H_0^T and the system output y. Visible stairstepping would probably appear if G had a first order rolloff. This partly explains why this system can have such a small cone radius: a system with visible stairstepping could not possibly be closely modelled by an LTI operator, so a larger radius would be needed to account for the error.

5.2 MIMO Single Rate Example

Figure 5.2-1 shows a MIMO sampled-data feedback system with multiplicative uncertainty and a single sampling period $T = .02$ second. All the operators shown — $G(s)$, $F(s)$, $H_0^T(s)$, $K(z)$, and ΔW — are 2×2 LTI transfer functions. $K(z)$ is a discrete time operator; the others are continuous time. $F(s)$ is a diagonal prefilter operator, and H_0^T in this example is a diagonal operator made of two ZOHs:

$$F(s) = \begin{bmatrix} \frac{2500}{s^2+70s+2500} & 0 \\ 0 & \frac{2500}{s^2+70s+2500} \end{bmatrix}$$

$$H_0^T(s) = \begin{bmatrix} \frac{1-e^{-sT}}{s} & 0 \\ 0 & \frac{1-e^{-sT}}{s} \end{bmatrix}$$

$G(s)$ is described by a 4-state state space model, and $K(z)$ by an 8-state model. Their state-space quadruples are shown in Figure 5.2-2. The multiplicative uncertainty block is factored into two parts W and Δ, where Δ is an uncertain full matrix satisfying $\bar{\sigma}(\Delta) < 1$ at all frequencies, and where W is a scalar weighting factor times an identity matrix. Δ and W are, as always, assumed to be stable.

This example is taken from [20]. $G(s)$ is a linearized model of the pitch axis motion of the HIMAT highly maneuverable aircraft. Its four states are forward velocity, angle of attack, pitch rate, and pitch attitude, and its two inputs are the two control surface positions. The two outputs are angle of attack and pitch attitude. In this example, we are not concerned with the physical system, however, but merely with testing the conservativeness of the new conic sector methods.

The MIMO digital compensator $K(z)$ was derived in [20]. First a stabilizing analog compensator was synthesized by the LQG loop transfer recovery technique of [3]. Then the analog compensator was discretized by using the Tustin (bilinear) transformation prewarped about 7 rad/sec, which is the approximate loop bandwidth. $K(z)$ has two poles at $z = 1$, but none outside the unit circle; $G(s)$ has two RHP poles.

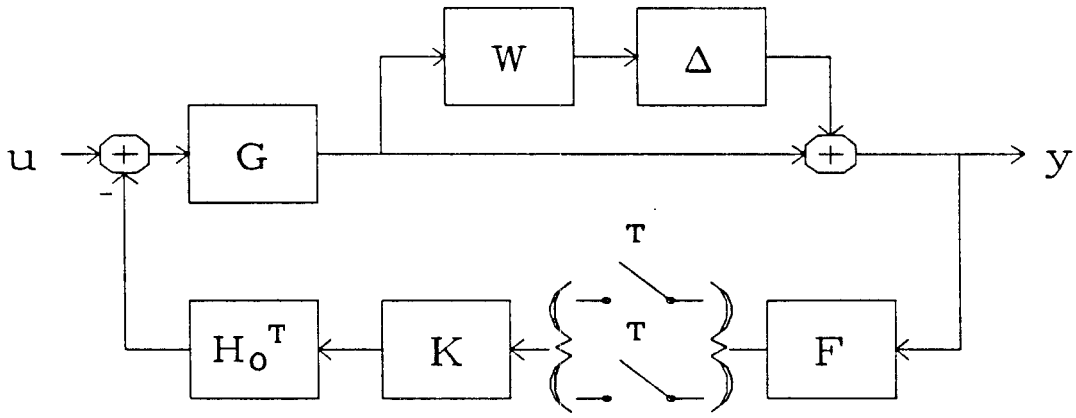

Fig. 5.2-1: A 2×2 MIMO feedback system with multiplicative uncertainty $1 + \Delta W$ at the output of the plant G.

```
Matrix : g
outputs 2   inputs 2   states 4   double precision

        x1          x2          x3          x4          u1          u2
  x1 -2.257e-02 -3.662e+01 -1.890e+01 -3.209e+01 -9.821e-01 -7.626e-01
  x2  9.000e-05 -1.900e+00  9.831e-01 -7.300e-03 -4.144e-01 -4.960e-03
  x3  1.233e-02  1.172e+01 -2.632e+00  0.000e+00 -7.781e+01  2.240e+01
  x4  0.000e+00  0.000e+00  1.000e+00  0.000e+00  0.000e+00  0.000e+00
  y1  0.000e+00  1.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00
  y2  0.000e+00  0.000e+00  0.000e+00  1.000e+00  0.000e+00  0.000e+00

Matrix : kdig
outputs 2   inputs 2   states 8   double precision

        x1          x2          x3          x4          x5          x6
  x1  9.995e-01  6.143e-02 -2.774e-01  3.663e+00  9.868e-02 -4.872e-02
  x2  1.787e-06  8.141e-01  1.218e-02 -8.341e-02 -8.945e-03  6.861e-04
  x3  9.415e-05 -3.243e+00  5.439e-01 -5.189e+00 -6.099e-01  1.884e-01
  x4  8.682e-07 -5.572e-02  1.447e-02  8.455e-01 -5.661e-03  1.777e-03
  x5  1.945e-04  7.350e+00  4.791e-01  2.809e+00 -1.278e-01  2.436e-01
  x6  6.134e-06  1.046e+01 -1.773e-01 -1.259e+01  2.362e-01  7.091e-01
  x7  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00
  x8  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00  0.000e+00
  y1 -1.501e-02 -7.209e+02 -4.929e+01 -6.171e+02  8.859e+01 -2.050e+01
  y2  9.874e-04 -4.316e+02  1.610e+01  7.873e+02 -2.050e+01  1.747e+01

        x7          x8          u1          u2
  x1 -3.266e-01 -3.311e+00 -5.560e-03 -6.410e-02
  x2  9.114e-02  2.257e-02  1.722e-03  3.553e-04
  x3 -1.043e-01 -6.662e-02 -5.234e-03 -5.774e-03
  x4  2.529e-02  1.055e-01  4.422e-04  2.007e-03
  x5 -8.183e-02  6.605e-01  3.481e-04  1.235e-02
  x6 -1.586e+00  5.678e-01 -2.628e-02  5.918e-03
  x7  1.000e+00  0.000e+00  2.003e-02  0.000e+00
  x8  0.000e+00  1.000e+00  0.000e+00  2.003e-02
  y1  0.000e+00  0.000e+00 -2.550e-01 -1.561e-01
  y2  0.000e+00  0.000e+00 -2.360e-01  3.437e-01
```

Fig. 5.2-2: State-space quadruples for the plant $G(s)$ and the digital compensator $K(z)$.

Before proceeding with conic sector analysis, we can provide a basis for comparison by using standard LTI techniques. Since $F(s)$ is a scalar times an identity matrix, it commutes with $1 + \Delta W$. If ΔW is assumed to be *constant* at all frequencies, then $1 + \Delta W$ also commutes with the sampler. In this case, we get the equivalent discrete time system of Figure 5.2-3. Provided that the nominal gain $-(FGH_0^T)^T K(I+(FGH_0^T)^T K)^{-1}$ seen by ΔW is stable, then the perturbed system is stable if

$$\bar{\sigma}(\Delta W)\bar{\sigma}\left((FGH_0^T)^T K(I+(FGH_0^T)^T K)^{-1}\right) \leq 1 \quad \text{at all frequencies.}$$

This can be written equivalently as

$$\bar{\sigma}(\Delta W) \leq \underline{\sigma}\left(((FGH_0^T)^T K(I+(FGH_0^T)^T K)^{-1})^{-1}\right) = \underline{\sigma}\left(I+((FGH_0^T)^T K)^{-1}\right)$$

This provides a robustness margin against this type of constant perturbation ΔW. Although somewhat unrealistic, since actual perturbations will certainly be frequency-dependent, this test does give some indication of the system's nearness to instability. The system will be stable for any constant ΔW satisfying

$$\bar{\sigma}(\Delta W) \leq \inf_{\omega \in R} \underline{\sigma}\left(I+\left((FGH_0^T)^T(e^{j\omega T})K(e^{j\omega T})\right)^{-1}\right)$$

The minimum singular value is plotted as the dashed line in Figure 5.2-8. Its minimum value is .487; this will provide a basis for comparison when we find robustness margins using conic sectors.

Figure 5.2-4 is a log Nyquist plot of $\det(I+(FGH_0^T)^T K)$ over frequencies $-\pi/T$ to π/T. Since $K(z)$ has poles at $z = 1$, the Nyquist contour has a semicircular notch of radius $.01T = .0004$ around $z = 1$. The notch maps into the large near-circle of the plot. If the notch radius were reduced, the circle would grow and the tips of the crescent shape would approach, but never touch, the real axis. Since the determinant makes two counter-clockwise encirclements of the origin, equal to the number of unstable poles of $(FGH_0^T)^T K$, the nominal discretized system is stable.

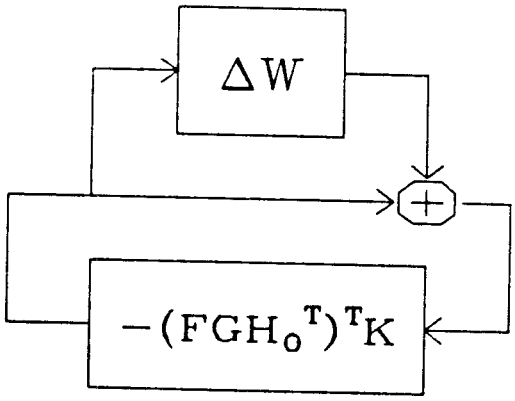

Fig. 5.2-3: Equivalent discrete-time system when ΔW is a constant.

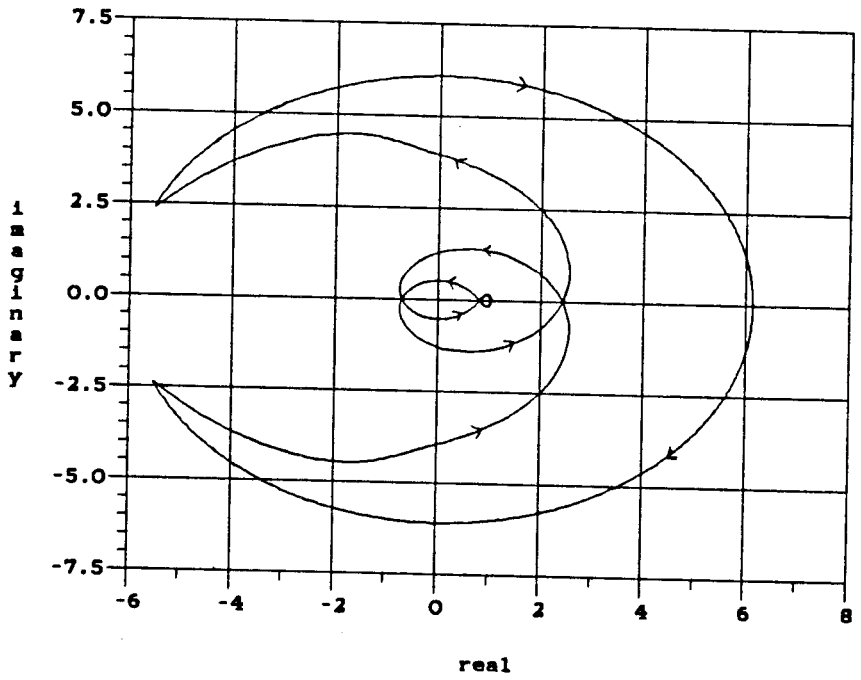

Fig. 5.2-4: MIMO log Nyquist plot of $\det(I + (FGH_0^T)^T K)$ over frequencies $-\pi/T$ to π/T. There are two CCW encirclements of 0, equal to the number of unstable poles of $(FGH_0^T)^T$, so the discretized system is stable.

Now we may proceed with the conic sector tests. Rearrange the system as in Figure 5.2-5, with the two samplers extracted above the system and the ΔW perturbation below. The 4×4 interconnection matrix M is given by

$$M = \begin{bmatrix} -FGH_0^T K & F \\ -GH_0^T K & -GH_0^T KF \end{bmatrix}$$

Next follow the usual analysis process. Sandwich the samplers between fictitious operators $A = \text{diag}(\alpha_1, \alpha_2)$ and $B = \text{diag}(\beta_1, \beta_2)$, then cone-bound them individually using Theorem 1. Compute $R = \text{diag}(\text{ Rad}(\alpha_1, \beta_1), \text{Rad}(\alpha_2, \beta_2))$ and its square root $R^{1/2}$. Absorb the normalized cone center $\hat{T} = \text{diag}(1/T, 1/T)$ into the system to produce the closed loop interconnection matrix $\tilde{M}$:

$$\begin{aligned} \tilde{M} &= \begin{bmatrix} \tilde{M}_{11} & \tilde{M}_{12} \\ \tilde{M}_{21} & \tilde{M}_{22} \end{bmatrix} \\ &= \begin{bmatrix} -FGH_0^T K(I + \frac{1}{T} FGH_0^T K)^{-1} & F(I + \frac{1}{T} GH_0^T KF)^{-1} \\ -GH_0^T K(I + \frac{1}{T} FGH_0^T K)^{-1} & -\frac{1}{T} GH_0^T KF(I + \frac{1}{T} GH_0^T KF)^{-1} \end{bmatrix} \end{aligned}$$

The system can now be redrawn as in Figure 5.2-6. $\hat{\Delta}$ is a PLTV operator inside $\text{Cone}(0, I)$, as usual. Scaling has been exploited for the analog Δ block, since $D_2 \Delta D_2^{-1} = \Delta$ for any LTI *scalar* D_2. D_2 must be scalar (or equivalently, a scalar times the identity matrix) because $\hat{\Delta}$ is now a full 2×2 matrix. The 4×4 operator $\hat{M}$ inside the dashed box is given by

$$\hat{M} = \begin{bmatrix} R^{1/2} A^{-1} \tilde{M}_{11} B^{-1} R^{1/2} & R^{1/2} A^{-1} \tilde{M}_{12} D_2^{-1} \\ D_2 W \tilde{M}_{21} B^{-1} R^{1/2} & W \tilde{M}_{22} \end{bmatrix}$$

Provided that the system meets the nominal stability test, it will be robustly stable if $\|\hat{M}\|_\infty < 1$.

Nominal stability. The system is nominally stable if $\tilde{M}_{11}$ is stable and if

$$\bar{\sigma}(R^{1/2} A^{-1} \tilde{M}_{11} B^{-1} R^{1/2}) < 1$$

at all frequencies. We can test the stability of $\tilde{M}_{11}$ by the MIMO Nyquist test. Figure 5.2-4 shows a log Nyquist plot of $\det(I + \frac{1}{T} FGH_0^T K)$. This same plot was

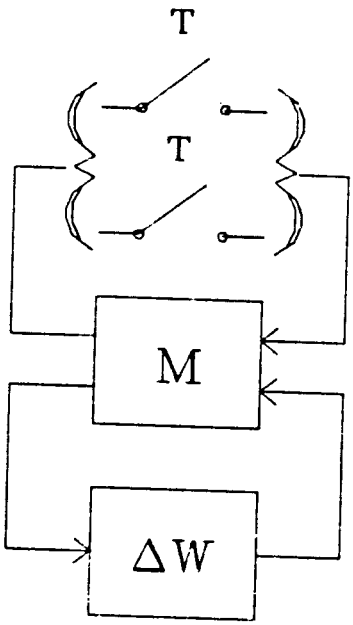

Fig. 5.2-5: Rearranged system for conic sector analysis.

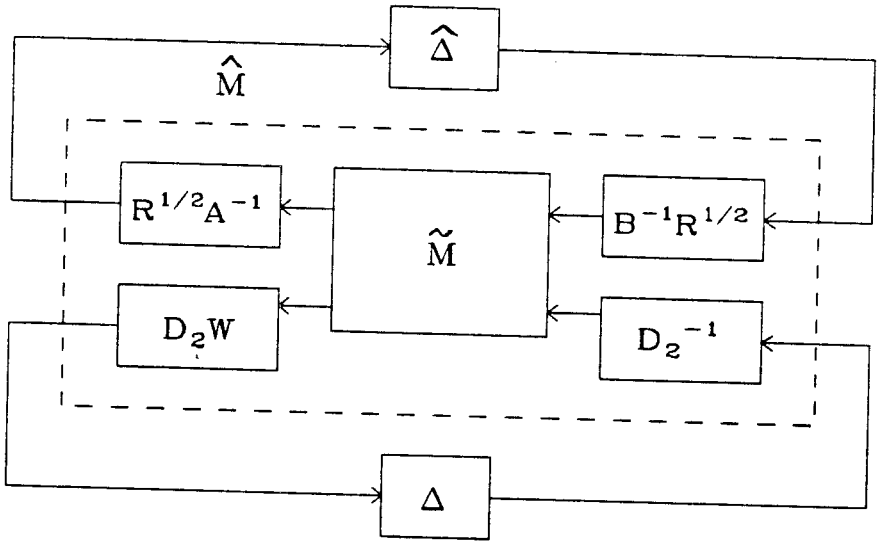

Fig. 5.2-6: Construction of $\hat{M}$ for robust stability test.

earlier identified as a log Nyquist plot of $\det(I + (FGH_0^T)^T K)$ for the discrete time stability test. Actually, it shows *both* functions superimposed. The two loci are so nearly identical that they are indistinguishable. This time, the Nyquist contour has a semicircular notch of radius .01 around $s = 0$. The earlier remarks about locus behavior as the notch shrinks still apply. Since the locus makes two counter-clockwise encirlements of the origin, equal to the number of unstable poles of $\frac{1}{T} FGH_0^T K$, $\tilde{M}_{11}$ is stable.

To compute the operators A, B, and R by the MIMO method of Section 3.7.1, find the diagonal LTI operator D which solves

$$\hat{\mu} \;=\; \inf_D \bar{\sigma}(D\tilde{M}_{11}D^{-1})$$

at each frequency. Then let $A = \hat{\mu}^{1/2} D^{-1}$ and $B = \hat{\mu}^{1/2} D$, where D has first been scaled to satisfy $\bar{\sigma}(D)\underline{\sigma}(D) = 1$. Since D is 2×2 in this case, this means that the two diagonal elements of D are reciprocals of each other; this in turn means that $\alpha_1 = \beta_2$ and $\alpha_2 = \beta_1$. Therefore both diagonal elements of the radius R are the same, and R is a scalar times identity in the single rate 2×2 case. While the inequality $\bar{\sigma}(R^{1/2} A^{-1} \tilde{M}_{11} B^{-1} R^{1/2}) \leq \bar{\sigma}(R)$ is true in general with these choices for A and B, in the single rate 2×2 case it holds with equality.

Figure 5.2-7 is a plot of $\bar{\sigma}(R) = \bar{\sigma}(R^{1/2} A^{-1} \tilde{M}_{11} B^{-1} R^{1/2})$ over frequency. Its maximum values is $\|R\|_\infty = .0123 \ll 1$; therefore the system is nominally stable.

Robust stability. To compute a frequency dependent robust stability margin, find the largest W at each frequency such that

$$\inf_{D_2} \bar{\sigma}(\hat{M}) \;=\; 1$$

Recall that D_2 and W are both scalars times identity matrices, so each has only one degree of freedom for the optimization problem. The solid line in Figure 5.2-8 is a plot of $\bar{\sigma}(W)$ derived in this way. Observe that, at low frequencies, it very

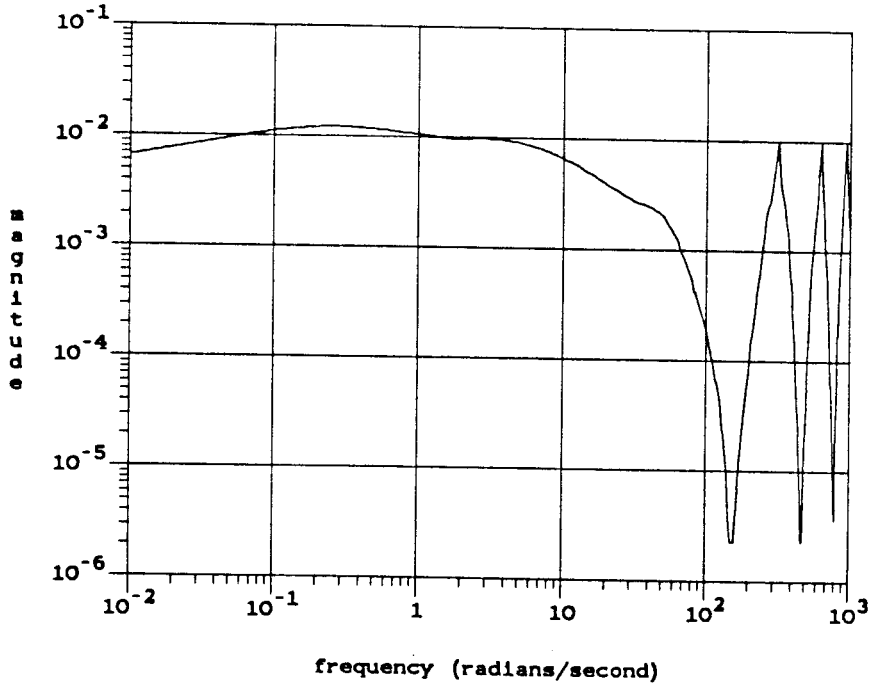

Fig. 5.2-7: Bode plot of $\bar{\sigma}(R) = \bar{\sigma}(R^{1/2}A^{-1}\tilde{M}_{11}B^{-1}R^{1/2})$. Since $\|R\|_\infty = .0123 \ll 1$, the system is nominally stable.

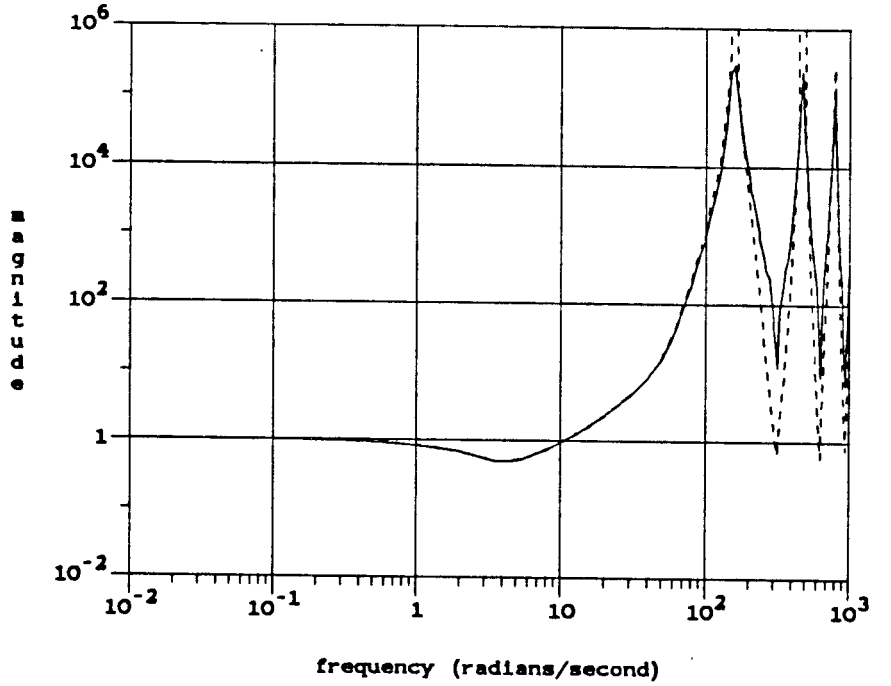

Fig. 5.2-8: Robust stability margins from conic sector (solid) and Kranc (dashed) analysis. The plots are nearly identical at low frequencies.

closely matches the dashed plot of $\underline{\sigma}\ (I + ((FGH_0^T)^T K)^{-1})$, which gave us a robustness margin for constant perturbations. The difference is that the solid line is a *frequency-dependent* robustness margin: any stable perturbation ΔW for which $\bar{\sigma}(\Delta W)$ lies below that curve at all frequencies will not destabilize the system. This gives us much more useful information than the discrete time test, for which only the minimum of the curve is a valid margin.

At the frequency where the dashed line reaches its minimum of .487, the solid line hits its minimum of .484. For this example, then, the conic sector method is conservative by only 0.7% in computing a bound for constant perturbations. In exchange for this tiny amount of conservativeness, conic sectors give a much more useful frequency-dependent robustness margin.

At high frequencies, $\bar{\sigma}(W)$ oscillates between large peaks of about 10^5 and deep minima of approximate size 10. If a larger robustness margin at high frequencies is desired, the choices for A and B may be manipulated to trade off larger high frequency margin for a slightly smaller margin at low frequencies. As $\omega \to \infty$, the A and B operators used above roll off approximately as $1/\omega^2$. By scaling A and B by the weighting function $|1 + \omega/100|$, we can make A and B roll off as $1/\omega$, still fast enough for $\mathrm{Rad}(\alpha_i, \beta_i)$ to converge. This will produce a larger radius R, but in exchange, it will cause 3 of the 4 block elements of $\hat{M}$ to roll off more quickly as $\omega \to \infty$. This allows the robustness margin W to be increased while still keeping $\bar{\sigma}(\hat{M}) = 1$.

When A and B are so modified, the new cone radius R has maximum gain $\|R\|_\infty = .0499$. This is four times larger than with the original A and B, but still less than 1. Therefore the system still passes the nominal stability test with this A, B. Now, when the maximum $\bar{\sigma}(W)$ is computed such that $\inf_{D_2} \bar{\sigma}(\hat{M}) = 1$, the result is the alternative robustness margin of Figure 5.2-9. This curve rolls up approximately as ω as $\omega \to \infty$, so it increases for high frequencies. At $\omega = 1000$, it

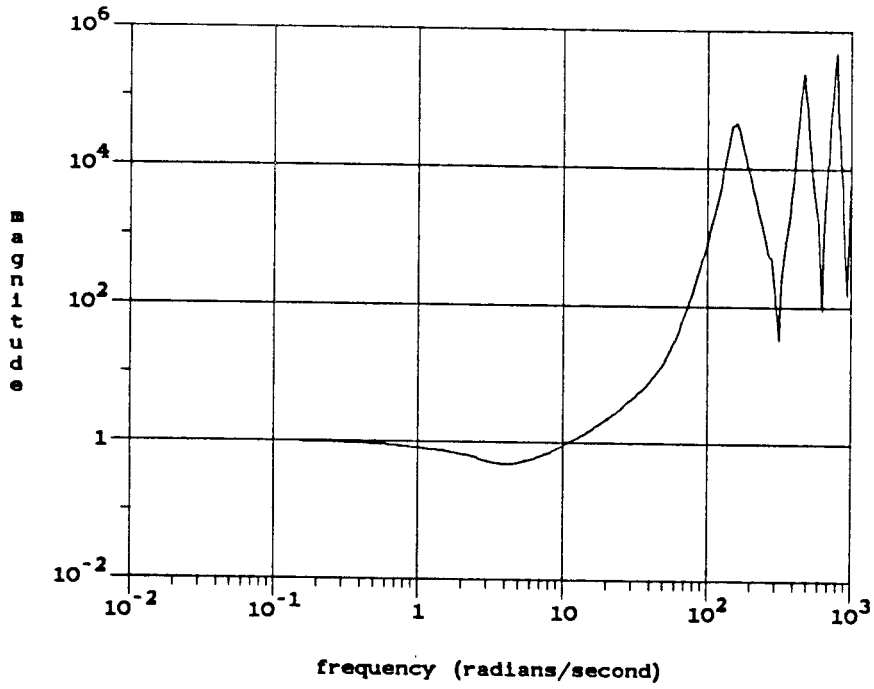

Fig. 5.2-9: Alternative conic sector robust stability margin. Margin is larger for larger ω and slightly smaller for small ω.

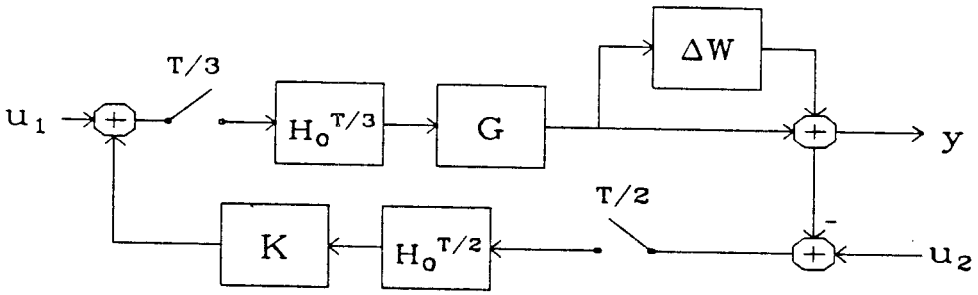

Fig. 5.3-1: A SISO multirate system.

is an order of magnitude larger than the margin of Figure 5.2-8. At low frequencies, it has a minimum value of .474, which is 2.1% smaller than the previous conic sector margin and 2.8% smaller than the absolute bound from discrete analysis.

The new conic sector methods of this thesis have proved to be very nonconservative for this MIMO example, which is based on a genuine control problem. The conic sector robustness margin is only 0.7% smaller than the actual (necessary and sufficient) margin derived by a standard LTI method. Moreover, the conic sector margin is much more useful than the standard result, since it gives a frequency dependent bound on allowable perturbations. Conic sector methods also allow tradeoffs between low frequency and high frequency model uncertainty.

5.3 Multirate Example

Figure 5.3-1 shows a single-loop SISO feedback system containing two samplers which run at different rates. The plant $G(s)$ is a lightly damped second order system, and the integrating compensator $K(s)$ provides a unity loop gain crossover frequency of 1 rad/sec. G is fed by a sampler of period $T/3$, and K by a sampler of period $T/2$, where $T = 0.1$ sec. Both samplers are followed by ZOHs of appropriate length. A stable multiplicative perturbation $1 + \Delta W$ acts on the output of G. As usual, Δ is an unknown perturbation satisfying $\|\Delta\|_\infty < 1$ and W is a weighting function. We will use robustness tests to find limits on W. The transfer functions F, G, and K are given by

$$F(s) = \frac{3600}{s^2 + 100s + 3600} \qquad G(s) = \frac{1}{s^2 + .4s + 1}$$

$$K(s) = \frac{312(s+1)^2}{s(s+30)^2}$$

Since the two sampling rates are related by a rational number, we may use the state-space Kranc operator techniques of Section 2.8 both to test for nominal stability

and to find a limit on *constant* perturbations ΔW. We will use the robustness margin from the Kranc method to test the conservativeness of the new conic sector methods for this example. Note, however, that the sampling rates do not have to be related by a rational number for the conic sector methods; this is a limitation only for the Kranc technique.

Kranc analysis. Divide the loop into two halves, operating from the $T/3$ sampler to the $T/2$ sampler and vice versa. The Kranc operators for each half are found by first discretizing at the short period $\gcd(T/2, T/3) = T/6$ with a short ZOH $H_0^{T/6}$. Let these two discretized operators be given by

$$(FGH_0^{T/6})^{T/6} \equiv \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right]$$

$$(FKH_0^{T/6})^{T/6} \equiv \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right]$$

Now to account for the longer lengths $T/2$ and $T/3$ of the actual ZOHs, we use the formula for forming Kranc operators with a ZOH:

$$(FG)^T_{Kranc} = \left[\begin{array}{c|ccc} A_1^6 & A_1^4 X_1 B_1 & A_1^2 X_1 B_1 & X_1 B_1 \\ \hline C_1 & D_1 & 0 & 0 \\ C_1 A_1^3 & C_1(A_1^2 + A_1 + I)B_1 & C_1 B_1 + D_1 & 0 \end{array}\right]$$

where $X_1 = A_1 + I$;

$$(FK)^T_{Kranc} = \left[\begin{array}{c|cc} A_2^6 & A_2^3 X_2 B_2 & X_2 B_2 \\ \hline C_2 & D_2 & 0 \\ C_2 A_2^2 & C_2(A_2 + I)B_2 + D_2 & 0 \\ C_2 A^4 & C_2 A_2 X_2 B_2 & C_2 B_2 + D_2 \end{array}\right]$$

where $X_2 = A_2^2 + A_2 + I$.

The state-space quadruple for the closed-loop Kranc operator of the entire system can now be formed using standard state-space system interconnection rules. The closed-loop poles are the eigenvalues of the state transition matrix for the closed-loop system. These poles were computed and are all inside the unit circle; therefore this system is closed-loop stable.

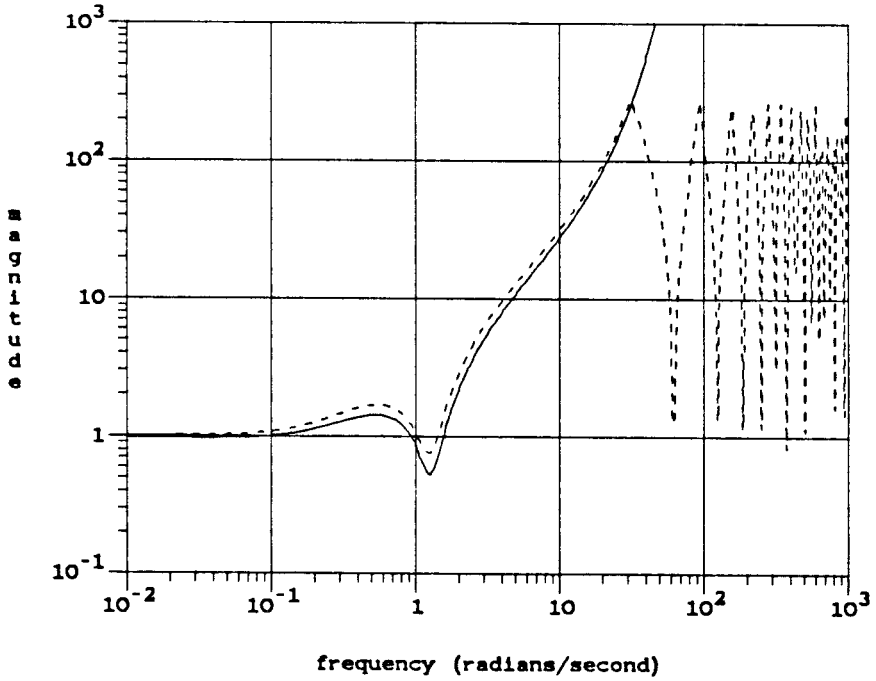

Fig. 5.3-2: Robust stability margins from Kranc (dashed) and conic sector (solid) analysis.

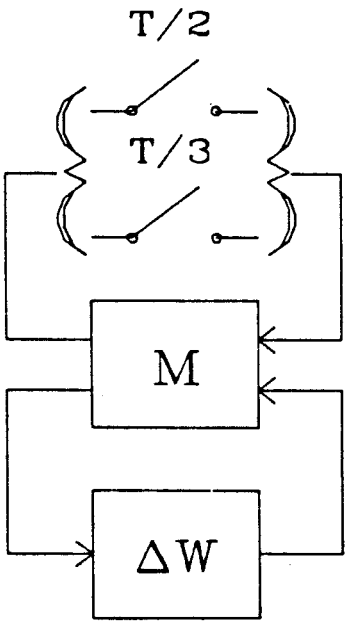

Fig. 5.3-3: The system rearranged for conic sector analysis.

Next we use the method from [20] described in Section 2.8.2 to find an upper bound on constant perturbations ΔW which maintain stability. Compute and plot

$$\left|\lambda_{\min}\left(I+\left((FG)^T_{Kranc}(FK)^T_{Kranc}\right)^{-1}\right)\right|$$

over frequency; the system is stable for any constant complex ΔW with magnitude less than this quantity at all frequencies. Figure 5.3-2 includes a plot of the minimum magnitude eigenvalue. The smallest value achieved by this function is .752 at $\omega = 1.26$. Therefore the system is stable for all constant $|\Delta W| < .752$.

Now we may proceed with the conic sector tests.

Nominal stability. Extract the two samplers and ΔW from the system as in Figure 5.3-3. The interconnection matrix M is given by

$$M = \begin{bmatrix} 0 & FKH_0^{T/2} & 0 \\ -FGH_0^{T/3} & 0 & -F \\ GH_0^{T/3} & 0 & 0 \end{bmatrix}$$

Form $\hat{T} = \mathrm{diag}(T/3, T/2)$ and lump $\hat{T}^{-1}$ together with M in the usual manner to produce the closed-loop cone center matrix $\tilde{M}$:

$$\tilde{M} = \begin{bmatrix} -\frac{2}{T}KH_0^{T/2}GH_0^{T/3} & KH_0^{T/2} & -\frac{2}{T}KH_0^{T/2} \\ -GH_0^{T/3} & -\frac{3}{T}GH_0^{T/3}KH_0^{T/2} & -1 \\ GH_0^{T/3} & \frac{3}{T}GH_0^{T/3}KH_0^{T/2} & -\frac{6}{T^2}GH_0^{T/3}KH_0^{T/2} \end{bmatrix} (1+L)^{-1}$$

where $L = \frac{6}{T^2}FKH_0^{T/2}FGH_0^{T/3}$ is the loop gain when the samplers are replaced by $3/T$ and $2/T$. Denote by $\tilde{M}_{\mathrm{ul}}$ the upper left 2×2 submatrix of $\tilde{M}$; this gives the interconnection of the two zero-center cone-bounded samplers. We can compute good choices for the sandwiching operators α_i, β_i by solving for

$$\hat{\mu} = \inf_D \bar{\sigma}(D\tilde{M}_{\mathrm{ul}}D^{-1})$$

at each frequency, where D ranges over all 2×2 diagonal matrices. Now construct $A = \mathrm{diag}(\alpha_1, \alpha_2) = \hat{\mu}^{1/2}D^{-1}$ and $B = \mathrm{diag}(\beta_1, \beta_2) = \hat{\mu}^{1/2}D$, where D has first been scaled to satisfy $\bar{\sigma}(D)\underline{\sigma}(D) = 1$ at every frequency. Sandwich the $T/3$ sampler

between α_1 and β_1 and the $T/2$ sampler between α_2 and β_2, and compute the cone radii $R_i = \mathrm{Rad}(\alpha_i, \beta_i)$. Note that since the sampling rates are different, the radii must be computed with different aliasing frequencies $\omega_s = 2\pi/T_i$. The use of different formulas for the radii means that $R_1 \neq R_2$, even though $\alpha_1 = \beta_2$ and $\alpha_2 = \beta_1$.

Figure 5.3-4 shows α_1 and α_2, and Figure 5.3-5 shows the two cone radii R_1 and R_2. Since $\max_i \|R_i\|_\infty$ is an upper bound for $\|R^{1/2}A^{-1}\tilde{M}_{\mathrm{ul}}B^{-1}R^{1/2}\|_\infty$, where $R = \mathrm{diag}(R_1, R_2)$, and since both R_1 and R_2 are less than 1, the system passes one of the two requirements for nominal stability. For completeness's sake, the quantity $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}_{\mathrm{ul}}B^{-1}R^{1/2})$ is shown in Figure 5.3-6; its maximum value is .336 at $\omega = 1.19$.

The other requirement is that $\tilde{M}_{\mathrm{ul}}$ be stable; this is equivalent to $(1 + L)^{-1}$ being stable. Figure 5.3-7 is a log Nyquist plot of L. Since L has a pole at $s = 0$, the D-contour has a notch of radius .01 around the origin; the notch maps into the large semicircle. The Nyquist locus makes no encirclements of the critical point -1. Since L has no unstable poles, this means that $\tilde{M}_{\mathrm{ul}}$ is stable, and this system is nominally stable.

Robust stability. To make the robust stability test, lump R, A^{-1}, B^{-1}, and W together with $\tilde{M}$ as shown in Figure 5.3-8. Also substitute $D_2 \Delta D_2^{-1}$ for Δ, and lump the scaling factors D_2 and D_2^{-1} with $\tilde{M}$ as well. The two samplers are now represented by the LTV operator $\hat{\Delta} = \mathrm{diag}(\hat{\Delta}_1, \hat{\Delta}_2)$, where $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are each inside $\mathrm{Cone}(0, I)$. Call the operator in the dashed box $\hat{M}$; it is given by

$$\hat{M} = \begin{bmatrix} R^{1/2}A^{-1} & 0 \\ 0 & D_2 W \end{bmatrix} \tilde{M} \begin{bmatrix} B^{-1}R^{1/2} & 0 \\ 0 & D_2^{-1} \end{bmatrix}$$

For any given uncertainty weight W, the system is robustly stable if $\|\hat{M}\|_\infty < 1$. D_2 is a free parameter at each frequency, since Δ is LTI and puts no periodicity constraint on D_2. We may compute a robustness margin by finding the largest W

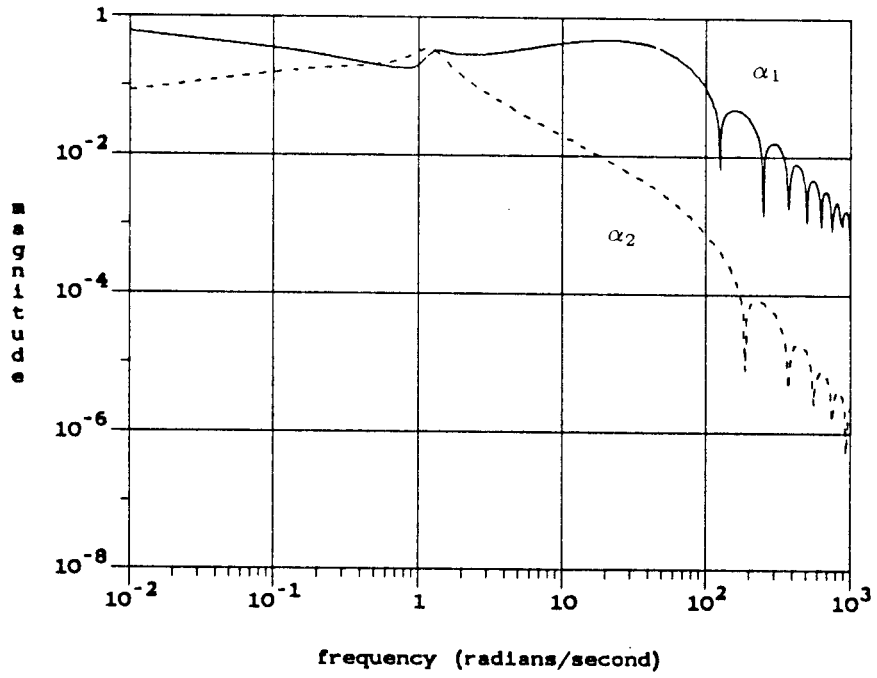

Fig. 5.3-4: Bode plots of the fictitious sandwiching operators $\alpha_1 = \beta_2$ and $\alpha_2 = \beta_1$.

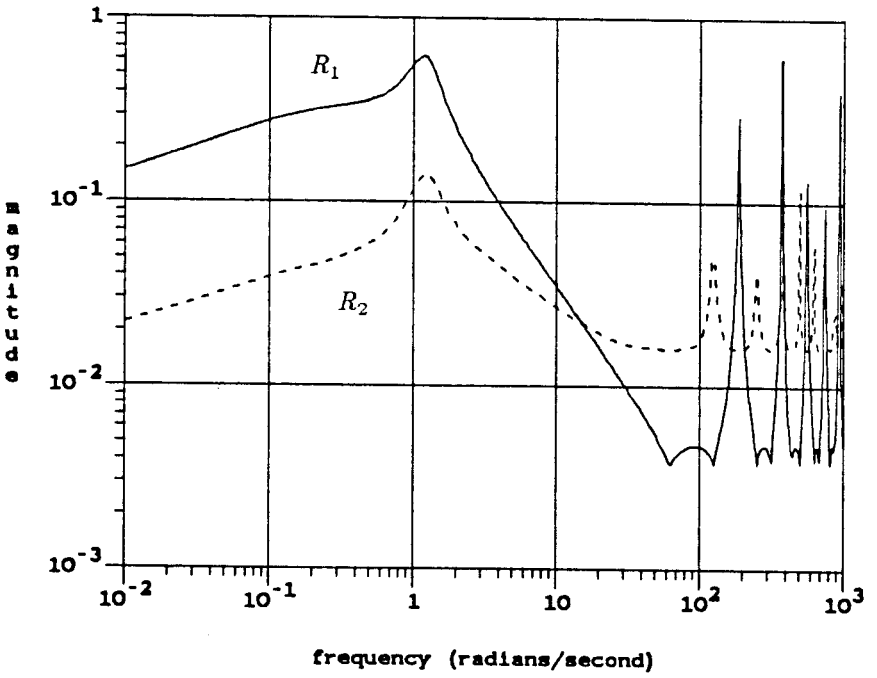

Fig. 5.3-5: SISO cone radii $R_1 = \mathrm{Rad}(\alpha_1, \beta_1)$, $R_2 = \mathrm{Rad}(\alpha_2, \beta_2)$. They are different because $T_1 \neq T_2$.

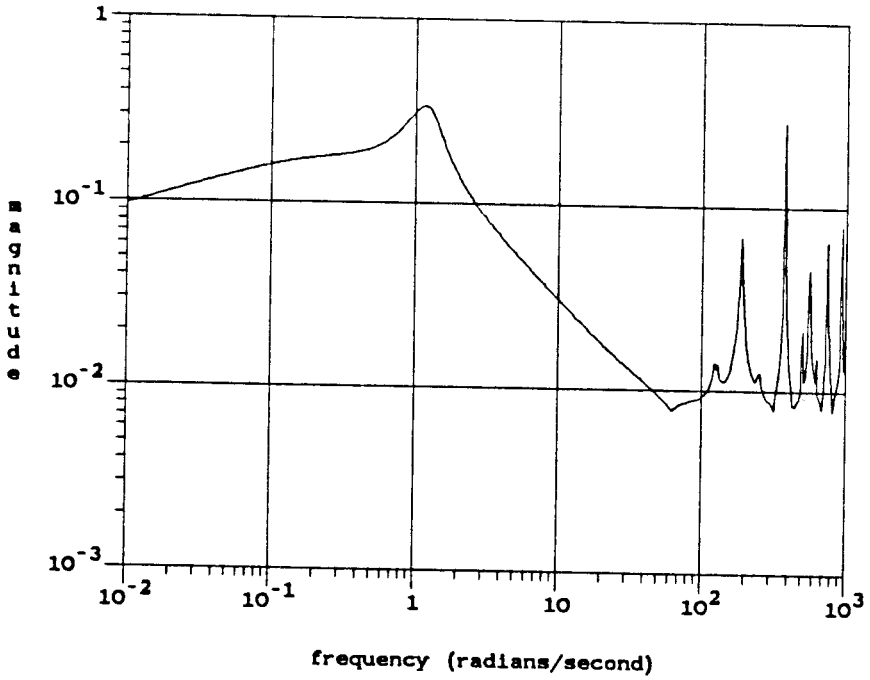

Fig. 5.3-6: Bode plot of $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}_{\rm ul}B^{-1}R^{1/2})$. Since the maximum amplitude is less than 1, the system passes part of the nominal stability test.

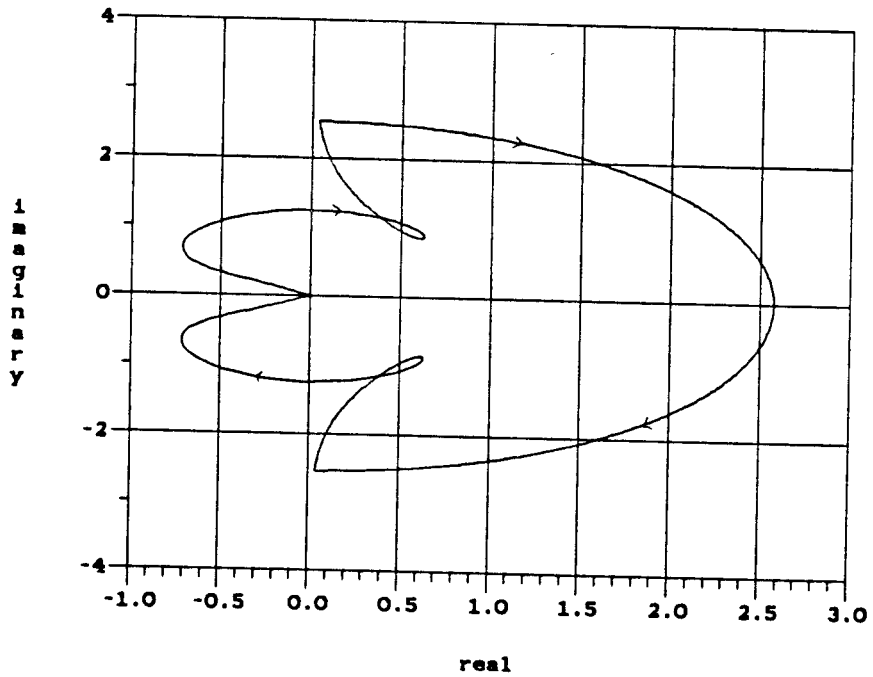

Fig. 5.3-7: Log Nyquist plot of $L = \frac{6}{T^2}FKH_0^{T/2}FGH_0^{T/3}$, the loop gain with samplers replaced by $1/T_i$. No encirclements of -1 implies nominal stability.

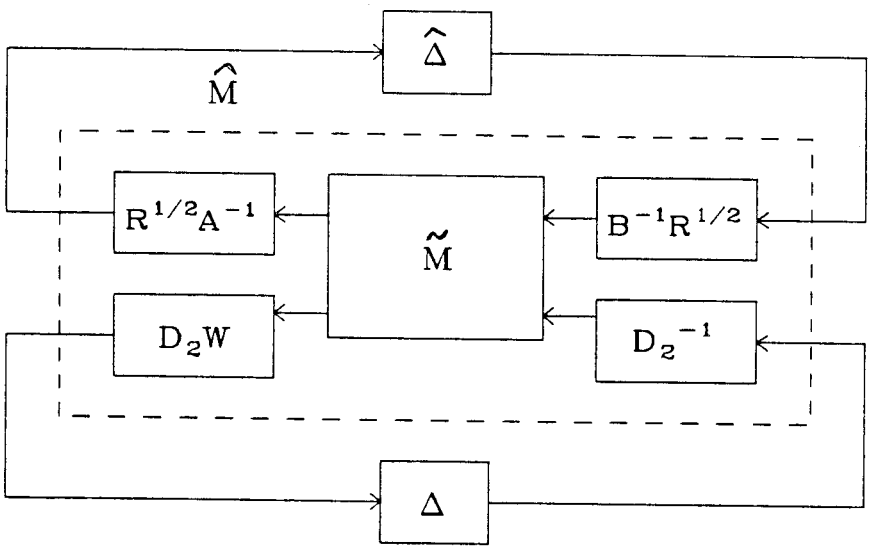

Fig. 5.3-8: Construction of $\hat{M}$ for the robust stability test.

at each frequency such that

$$\inf_{D_2 \in R} \bar{\sigma}(\hat{M}) = 1$$

Figure 5.3-2 shows the result of such a calculation. $W_{\max}$ is the robustness margin. The system is robustly stable for any stable perturbation ΔW satisfying $|\Delta W| < W_{\max}$ at all frequencies. The Kranc robustness margin

$$|\lambda_{\min}(I + ((FG)^T_{Kranc}(FK)^T_{Kranc})^{-1})|$$

is shown on the same plot. At $\omega = 1.26$, where the Kranc margin reaches its minimum value of .752, $W_{\max} = .528$. This means that the conic sector robustness margin is conservative at this frequency by 30%. However, the conic sector margin is still more useful, since it is frequency dependent. For example, the plot of $W_{\max}$ shows that at $\omega = 10$, a perturbation with magnitude nearly 30 is acceptable; the LTI Kranc margin makes no guarantees if $|\Delta W|$ is greater than .752 at any frequency.

The conservativeness of the new method for this example does not imply that conic sector methods are always conservative for multirate systems. In this example the conservativeness is mostly due to the high gain of K above the crossover frequency. From $\omega = 1$ to $\omega = 30$, $|K|$ increases at a 20 dB/decade slope. As a result, α_1 is relatively large at high frequencies, causing R to be large in turn. In general, conic sector robustness margins are the most conservative at frequencies where R is largest; as R approaches 1, the margin approaches zero.

Chapter Six — Synthesis of Digital Controllers

The previous chapters have dealt with the analysis of sampled-data feedback systems after the systems have been designed. In this chapter, we will touch on the design of digital controllers to achieve desired closed loop performance.

The literature on optimal control system synthesis is vast and beyond the scope of this thesis. In this chapter, we will deal only with the synthesis of digital controllers which approximately match the response of given analog controllers. The analog controllers may be arrived at by any design method, such as μ-synthesis, H_∞-synthesis, LQG, singular value loop shaping, pole placement, etc. Once a "target" analog controller is decided on, a stabilizing digital controller can be found for which the cone center frequency response matches that of the target as closely as possible.

To be more specific, suppose the target is a SISO controller $K_a(s)$. We wish to replace it with the hybrid controller in Fig. 6.1-1, which includes a prefilter $F(s)$, a sampler of period T, a digital filter $K_d(z)$, and a hold element $H(s)$. When the sampler is cone-bounded in the manner of the previous chapters, and the resulting cone center is lumped with the system, the effect is the same as replacing the sampler with $1/T$. The cone center frequency response of the controller, then, is $\frac{1}{T}HK_dF$. We would like to synthesize $K_d(z)$ so that $\frac{1}{T}HK_dF \approx K_a$ at frequencies inside the loop bandwidth; then the hybrid controller will have much the same effect on the system as the original analog target. Of course, it is impossible to match K_a at all frequencies; since K_d is periodic in ω, if we match closely from $\omega = 0$ to π/T, then

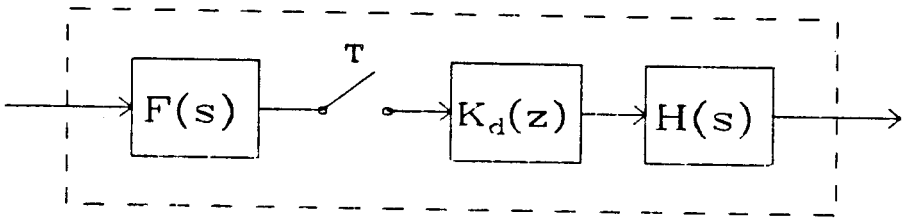

Fig. 6.1-1: A SISO hybrid controller.

$\frac{1}{T}HK_dF$ is predetermined at all higher frequencies. However, since the controller has little effect on the system for ω above the loop bandwidth, the high-frequency error is not very important.

Sections 6.1 through 6.3 outline a technique based on the complex cepstrum which chooses $K_d(z)$ to closely match the target TK_a/HF from $\omega = 0$ to π/T. The technique gives us control over the number of stable and unstable poles of K_d and the number of minimum phase and maximum phase zeros. It matches both phase and the log of magnitude, so it matches equally well at frequencies where the target magnitude is small or large. It has applications beyond control system synthesis, such as matching rational transfer functions to measured frequency response data.

Section 6.4 gives an example of matching a target analog transfer function with this method. Section 6.5 presents an example of digital controller synthesis.

6.1 The Complex Cepstrum

The complex cepstrum is a discrete time sequence which is derived from an original time sequence by a nonlinear three-step process [14]. First evaluate the z-transform of the original sequence $x(n)$ on the unit circle $|z| = 1$. This is simply the frequency response if the sequence is the impulse response of a system; call it $X(e^{j\omega T})$. Second, compute the complex logarithm of $X(e^{j\omega T})$; recall that $\mathrm{Re}[\log(X(e^{j\omega T}))]$ is the log of the magnitude of $X(e^{j\omega T})$, and $\mathrm{Im}[\log(X(e^{j\omega T}))]$ is its phase in radians. Third, find the Fourier series components for $\log[X(ej\omega T)]$ and arrange them as a time sequence:

$$\hat{x}(n) = \frac{T}{2\pi}\int_0^{2\pi/T} e^{j\omega nT}\log[X(e^{j\omega T})]d\omega$$

The time sequence $\hat{x}(n)$ is the complex cepstrum of $x(n)$.

The complex cepstrum was first introduced in [1] in 1963. Since then it has been used mostly within the digital signal processing community for such applications as

echo removal and finding pole-zero models of the human vocal tract [14]. Apparently, in most applications only the real part (log magnitude) of $\log(X(e^{j\omega T}))$ is of interest. In this chapter, the imaginary part is of equal importance, since we wish to match the phase as well as the log magnitude of the target transfer function.

The complex cepstrum is very well suited to computer calculation, with one caveat. Rather than compute the integral in the third step above, it is much easier to first sample $\log(X(e^{j\omega T}))$ on an evenly spaced grid of frequency points, and then to compute $\hat{x}(n)$ with the Fast Fourier Transform (FFT). The well-known efficiency of the FFT algorithm speeds the calculation tremendously. Since sampling in the frequency domain introduces aliasing in the time domain, however, it is important to sample the unit circle on a sufficiently large number of points. For a typical transfer function $X(e^{j\omega T})$, $512 = 2^9$ points are more than enough to make aliasing effects negligible.

Care must be taken in computing the phase $\mathrm{Im}[\log(X(e^{j\omega T}))]$. Since $\log(s)$ is not a one-to-one mapping, the correct branch of the function must be chosen to avoid phase discontinuities. Standard computer inverse tangent routines return values in the range $-\pi$ to π; sometimes multiples of 2π must be added or subtracted from this to avoid phase discontinuity. If the frequency point spacing is sufficiently small, this "phase unwrapping" problem can be handled simply in software.

Note that $\hat{x}(n)$ is always real if the original sequence $x(n)$ is real.

6.2 Complex Cepstrum of a Rational Function in z

If a sequence's z-transform $X(z)$ is a rational function in z, its complex cepstrum

can be written in closed form [14]. Let

$$X(z) = \frac{A\prod_{k=1}^{N_a}(1-a_kz^{-1})\prod_{k=1}^{N_b}(1-b_kz)}{\prod_{k=1}^{N_c}(1-c_kz^{-1})\prod_{k=1}^{N_d}(1-d_kz)}$$

where $|a_k|$, $|b_k|$, $|c_k|$, and $|d_k|$ are all less than 1, and $A > 0$. Then the coefficients a_k and c_k are zeros and poles, respectively, that are inside the unit circle. The zeros and poles outside the unit circle are given by $1/b_k$ and $1/d_k$. The log of $X(z)$ is

$$\begin{aligned}\log(X(z)) &\equiv \hat{X}(z)\\ &= \log(A) + \sum_{k=1}^{N_a}\log(1-a_kz^{-1}) + \sum_{k=1}^{N_b}\log(1-b_kz)\\ &\quad - \sum_{k=1}^{N_c}\log(1-c_kz^{-1}) - \sum_{k=1}^{N_d}\log(1-d_kz)\end{aligned}$$

Finding the Fourier series for $\hat{x}(z)$ evaluated on the unit circle is equivalent to finding the time sequence $\hat{x}(n)$ which has $\hat{X}(z)$ as its z-transform. We can do this in closed form by using these two power series expansions:

$$\log(1-\alpha z^{-1}) = -\sum_{n=1}^{\infty}\frac{\alpha^n}{n}z^{-n} \quad \text{for } |z| > |\alpha|$$

$$\log(1-\beta z) = -\sum_{n=1}^{\infty}\frac{\beta^n}{n}z^{n} \quad \text{for } |z| < |\beta^{-1}|$$

For the above assumptions and for z on the unit circle,

$$\max_k\{|a_k|,|c_k|\} < |z| = 1 < \min_k\{|b_k^{-1}|,|d_k^{-1}|\}$$

implies that the power series expansions converge on the unit circle when applied to $\hat{X}(z)$. Using the expansions to express $\hat{X}(z)$ as an infinite geometric series in z lets us read off the time sequence $\hat{x}(n)$ directly:

$$\hat{x}(n) = \begin{cases}\log|A| & n = 0\\ -\sum_{k=1}^{N_a}\frac{a_k^n}{n} + \sum_{k=1}^{N_c}\frac{c_k^n}{n} & n > 0\\ \sum_{k=1}^{N_b}\frac{b_k^{-n}}{n} - \sum_{k=1}^{N_d}\frac{d_k^{-n}}{n} & n < 0\end{cases}$$

This formula gives the entire complex cepstrum for $X(z)$ directly in terms of its poles and zeros. Note that a_k and c_k, the minimum phase zeros and poles, appear in the cepstrum only for $n > 0$. Likewise the maximum phase (outside unit circle) zeros and poles $1/b_k$ and $1/d_k$ appear only for $n < 0$. This means that the complex cepstrum maps minimum phase transfer functions into *causal* time sequences, and it maps maximum phase transfer functions into *anti-causal* time sequences. If a transfer function is a product of minimum phase and maximum phase functions, the cepstrum maps them separately into the $n > 0$ and $n < 0$ regions.

This immediately suggests a practical application. Suppose only the magnitude of a minimum phase discrete-time transfer function $G(e^{j\omega T})$ is known, and it is desired to synthesize its phase. Computing the complex cepstrum of the squared magnitude $|G|^2 = G(e^{j\omega T})G(e^{-j\omega T})$ maps the original function $G(e^{j\omega T})$ into a causal sequence and maps its maximum phase complex conjugate $G(e^{-j\omega T})$ into an anti-causal one. The causal portion can be extracted by setting $\hat{x}(n) = 0$ for $n < 0$ and halving the $\hat{x}(0)$ term. Then run the operation backward, performing a forward FFT on $\hat{x}(n)$ to get $\hat{X}(e^{j\omega T})$, then finding $\exp(\hat{X}(e^{j\omega T})) = X(e^{j\omega T})$ to produce a transfer function with the original magnitude and with phase response corresponding to a minimum phase function.

The form of the expression for $\hat{x}(n)$ gives us individual control over the number of minimum phase and maximum phase poles and zeros in the cepstrum-matching optimization problem of the next section.

6.3 Optimal Matching of the Complex Cepstrum

Given a complex cepstrum sequence $\hat{x}(n)$, we can set up a nonlinear optimization problem to find the best fit between $\hat{x}(n)$ and the cepstrum of a rational function with specified numbers of poles and zeros a_k, b_k^{-1}, c_k, d_k^{-1}. We will choose, as a cost

function J, the total squared error between $\hat{x}(n)$ and its best fit; this is the squared l_2-norm of the error. But since the squared l_2-norm of a Fourier series equals the mean-square level of the original periodic waveform, a fit which minimizes this J also minimizes the mean square error for $\hat{X}(e^{j\omega T})$ in the frequency domain. The result is a least-squares fit to the log of the target transfer function.

6.3.1 The Cost Function J and its Gradient

The scaling factor A can be found immediately by the equation $\hat{x}(0) = \log(|A|)$. Since the minimum phase and maximum phase portions of the transfer function are partitioned into the $n > 0$ and $n < 0$ regions of $\hat{x}(n)$, we can set up separate optimization problems for the minimum phase poles and zeros a_k, c_k and the maximum phase b_k^{-1}, d_k^{-1}. For the problem involving a_k, c_k, define the cost function by

$$J = \sum_{n=1}^{\infty} |\hat{e}(n)|^2 \quad \text{where } \hat{e}(n) = \hat{x}(n) + \sum_{k=1}^{N_a} \frac{a_k^n}{n} - \sum_{k=1}^{N_c} \frac{c_k^n}{n}$$

Here $\hat{x}(n)$ is the complex cepstrum of the target function, and the terms involving a_k and c_k give the complex cepstrum of the best fit. Since the transfer functions we are interested in always have real-valued impulse responses, $\hat{x}(n)$ can be assumed real. By restricting all complex a_k and c_k to complex conjugate pairs, we can insure that $\hat{e}(n)$ is real and drop the absolute value sign above. With these assumptions, the gradient is readily computed:

$$\frac{\partial J}{\partial a_k} = 2 \sum_{n=1}^{\infty} \hat{e}(n) a_k^{n-1}, \qquad \frac{\partial J}{\partial c_k} = -2 \sum_{n=1}^{\infty} \hat{e}(n) c_k^{n-1}$$

Now the reason is apparent for performing optimization in the time domain, on $\hat{x}(n)$, rather than directly in the frequency domain on $\hat{X}(e^{j\omega T}) = \log(X(e^{j\omega T}))$. The cost function and gradient both are given by simple formulas which can be efficiently computed. This advantage is in addition to the complex cepstrum's clean separation

of minimum phase and maximum phase factors. Furthermore, since $\hat{e}(n)$ and J are real, the computation can be done mostly with real arithmetic if a suitable real parameterization of poles and zeros is used.

Of course, the true infinite series cannot be computed; the series must be truncated. Since $\hat{x}(n)$ decays rapidly as $n \to \infty$, the series can be truncated to a relatively small number of terms without sacrificing accuracy.

For the anti-causal part of $\hat{x}(n)$, define the cost function J by

$$J = \sum_{n=-1}^{-\infty} |\hat{e}(n)|^2 \quad \text{where } \hat{e}(n) = \hat{x}(n) - \sum_{k=1}^{N_b} \frac{b_k^{-n}}{n} + \sum_{k=1}^{N_d} \frac{d_k^{-n}}{n}$$

This can be rewritten as

$$J = \sum_{n=1}^{\infty} |\hat{e}(-n)|^2$$

with $\hat{e}(n)$ the same as before. Now the gradient is

$$\frac{\partial J}{\partial b_k} = 2 \sum_{n=1}^{\infty} \hat{e}(-n) b_k^{n-1} \qquad \frac{\partial J}{\partial d_k} = -2 \sum_{n=1}^{\infty} \hat{e}(-n) d_k^{n-1}$$

6.3.2 Real Parameterization of Poles and Zeros

Since all complex poles and zeros come in conjugate pairs, it is possible to parameterize N poles or zeros with N real numbers. For instance, parameterize the set of minimum phase zeros $\{a_k\}$ by the N_a real numbers $\{\tilde{a}_k\}$, where

$$a_k = \begin{cases} \tilde{a}_k + \tilde{a}_{k+1} & \text{if } \tilde{a}_{k+1} \geq 0 \\ \tilde{a}_k + j\tilde{a}_{k+1} & \text{if } \tilde{a}_{k+1} < 0 \end{cases} \quad \text{and}$$

$$a_{k+1} = \begin{cases} \tilde{a}_k - \tilde{a}_{k+1} & \text{if } \tilde{a}_{k+1} \geq 0 \\ \tilde{a}_k - j\tilde{a}_{k+1} & \text{if } \tilde{a}_{k+1} < 0 \end{cases}$$

for all *odd* $k < N_a$. If N_a is odd, let the last zero be assumed real and let $a_{N_a} = \tilde{a}_{N_a}$. This parameterization causes no loss of generality, since it allows as many as N_a non-real zeros and as few as none. The sign of each even-indexed parameter $\tilde{a}_{k+1}$ determines whether that pair of zeros is real or complex. If $\tilde{a}_{k+1} = 0$, a double

zero is produced. Let all b_k, c_k, and d_k be parameterized in the same way. Now recompute the gradients:

$$\frac{\partial J}{\partial \tilde{a}_k} = \left(\frac{\partial J}{\partial a_k}\frac{\partial a_k}{\partial \tilde{a}_k}\right) + \left(\frac{\partial J}{\partial a_{k+1}}\frac{\partial a_{k+1}}{\partial \tilde{a}_k}\right) = 2\sum_{n=1}^{\infty} \hat{e}(n)[a_k^{n-1} + a_{k+1}^{n-1}] \quad \text{and}$$

$$\frac{\partial J}{\partial \tilde{a}_{k+1}} = \left(\frac{\partial J}{\partial a_k}\frac{\partial a_k}{\partial \tilde{a}_{k+1}}\right) + \left(\frac{\partial J}{\partial a_{k+1}}\frac{\partial a_{k+1}}{\partial \tilde{a}_{k+1}}\right)$$

$$= \begin{cases} 2\sum_{n=1}^{\infty} \hat{e}(n)[a_k^{n-1} - a_{k+1}^{n-1}] & \text{if } \tilde{a}_{k+1} \geq 0 \\ 2j\sum_{n=1}^{\infty} \hat{e}(n)[a_k^{n-1} - a_{k+1}^{n-1}] & \text{if } \tilde{a}_{k+1} < 0 \end{cases}$$

for all odd $k < N_a$. If N_a is odd, $\partial J/\partial a_{N_a}$ is defined as before. Note that the gradient is always real, even when $\tilde{a}_{k+1} < 0$, since in that case $[a_k^{n-1} - a_{k+1}^{n-1}]$ is purely imaginary and is made real by the factor $2j$ in the last equation.

The other gradients are given by the same formula when b_k, c_k, or d_k are substituted for a_k, except that the sign of the gradient is reversed for c_k and d_k. Also, $\hat{e}(n)$ must be replaced by $\hat{e}(-n)$ for the gradients with b_k and d_k.

Complex arithmetic is still required to compute the terms a_k^n. However, this parameterization reduces the optimization problem to one involving, e.g., $N = N_a + N_c$ real variables instead of twice that number. This speeds the calculation greatly and automatically insures that the poles and zeros appear in complex conjugate pairs.

With the above formulas for the cost function J and its gradient with respect to all a_k, b_k, c_k, and d_k, any gradient-based nonlinear optimization technique may be used to find the best fit to a given complex cepstrum. The author has written a FORTRAN program which uses the Davidon-Fletcher-Powell (DFP) method to solve this problem [9]. The user enters the desired numbers of poles and zeros N_a, N_b, N_c, and N_d, and also enters a convergence test limit ϵ. The program iteratively searches for the minimum J until the relative change in J from one iteration to

the next is less than ϵ. Then it reverses the real parameterization and prints the complex pole and zero locations of the best fit.

6.3.3 Applications

The method above finds a best fit to a discrete-time transfer function evaluated on the unit circle. It can also be used to find a best fit to an analog transfer function, evaluated on the $j\omega$-axis. This can be done by mapping the target analog function onto the unit circle using the prewarped Tustin transformation:

$$s = \alpha\frac{z-1}{z+1} \qquad\qquad z = \frac{\alpha+s}{\alpha-s}$$

where $\alpha > 0$. This maps each analog frequency $s = j\omega$ into a point on the unit circle $e^{j\omega'}$ where $\omega' = 2\tan^{-1}(\omega/\alpha)$. The analog frequency $j\omega_0 = j\alpha$ is mapped into $z = e^{j\pi/2}$. By choosing α to map the most critical frequency range of the analog target into the region around $e^{j\pi/2}$, one can insure the best possible fit in that frequency range.

Usually, the analog target function is not available as a rational function in s, however; instead it is given as a set of frequency points. This situation is easily handled by applying the mapping $\omega' = 2\tan^{-1}(\omega/\alpha)$ to each frequency point. The resulting set of points can be sampled on an evenly spaced grid of, say, 512 points from 0 to 2π; grid points which lie between the given points can be estimated by linear, cubic spline, or any other interpolation scheme. Then $\log(H(e^{j\omega'}))$ can be computed for each grid point, the complex cepstrum computed by FFT, and the optimization problem solved.

The example of the next section starts with an analog target function and uses this prewarped Tustin mapping to convert it to a discrete-time function. Since the optimization problem returns poles and zeros in the z-plane, the last step is to convert them back into the analog s-plane with the inverse Tustin mapping: a pole

or zero at z_0 maps to a pole or zero at $s_0 = \alpha(z_0 - 1)/(z_0 + 1)$.

This technique is useful for fitting rational functions in s to a set of measured frequency response points of an analog system. Another application is SISO model reduction. If a high-order rational function in s or z is given, and it is desired to approximate it with a lower-order function, then the frequency response of the target function (or its Tustin mapping) can simply be evaluated on a unit circle grid of points. After the cepstrum is computed, one can experiment with various numbers of poles and zeros (minimum and maximum phase) to find the lowest order rational function that gives an acceptable fit.

Once a state-space representation of the best fit is set up, standard numerical integration methods can be used to perform I/O simulation, such as finding the step response. In this way accurate predictions of time domain response can be made from a set of frequency response points. This was done for the example system of Section 5.1. Of course, if the frequency response of a discrete-time system is known, its impulse response can be computed directly via the FFT. In the analog case, however, directly sampling the frequency response and performing an FFT can lead to inaccuracies due to aliasing in the time domain. Fitting the frequency response first to a rational function in s sidesteps this problem.

6.4 Example: Matching an Analog Transfer Function

Section 5.1 used a system's closed-loop cone center transfer function

$$C(j\omega) \;=\; G(j\omega)\left(1 + \frac{1}{T}H_0^T(j\omega)K(e^{j\omega T})F(j\omega)G(j\omega)\right)^{-1}$$

to approximate the closed-loop step response. An approximating state-space quadruple was found using the complex cepstrum matching method and then the step response was generated by standard numerical integration techniques. This section describes in detail the steps involved.

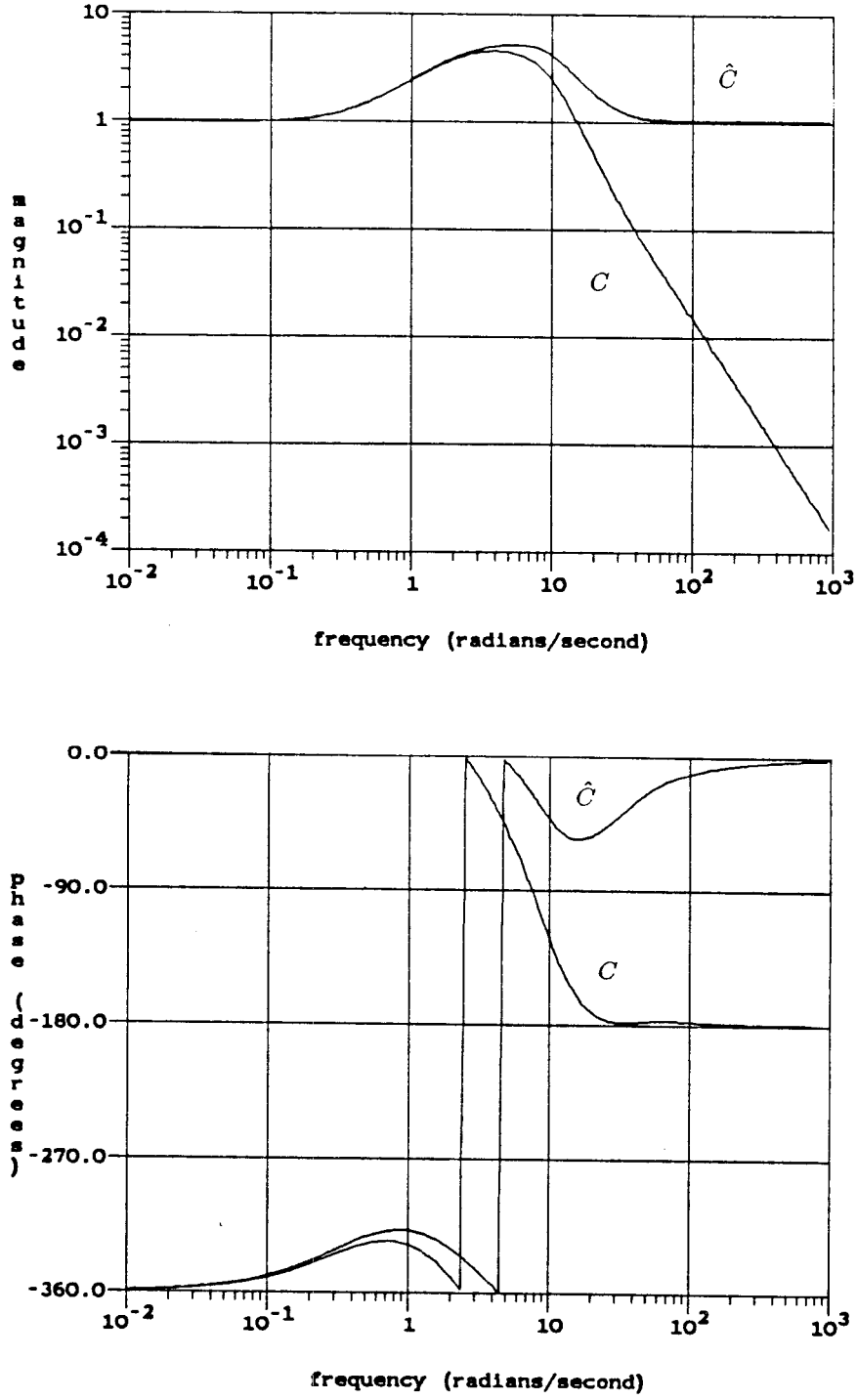

Fig. 6.4-1: Magnitude and phase of original and prescaled target functions $C(j\omega)$ and $\hat{C}(j\omega)$.

Figure 6.4-1 shows the magnitude and phase of $C(j\omega)$. C rolls off at a -40 dB/decade slope as $\omega \to \infty$. Since computing the complex cepstrum involves finding the log of $C(j\omega)$, it is not permissible for $|C(j\omega)|$ to go to zero; this would cause $\mathrm{Re}[\log(C(j\omega))]$ to go to $-\infty$. To avoid this problem, prescale C with two zeros at $s = -12$:

$$\hat{C}(j\omega) = C(j\omega)\left(\frac{s+12}{12}\right)^2$$

This prescaling cancels the high-frequency rolloff of C, so that $\hat{C}(j\omega)$ is nonzero as $\omega \to \infty$. Figure 6.4-1 shows $\hat{C}(j\omega)$. The new zeros also change the net phase shift as ω goes from 0 to ∞; $\hat{C}$ has a net change of 0°, while C changes by 180°. It is important that the net phase shift be zero, so that no discontinuities appear in $\hat{X}(e^{j\omega'}) = \log X(e^{j\omega'})$ as $e^{j\omega'}$ ranges around the unit circle.

The next step is to use the prewarped Tustin transformation to map $\hat{C}(j\omega)$ onto the unit circle. We will set the prewarping parameter $\alpha = 12$. Since $\hat{C}(j\omega)$ is given to us as a set of frequency points, the transformation is easily accomplished by replacing each frequency ω with $\omega' = 2\tan^{-1}(\omega/\alpha)$. Now the discrete-time equivalent of $\hat{C}$ is $X(e^{j\omega'}) = \hat{C}(j\omega)\ \forall\, \omega \in R$. Note that the zeros at $s = -12 = -\alpha$ map to $z = (\alpha + s)/(\alpha - s) = 0$. The resulting unevenly spaced set of points for $X(e^{j\omega'})$ was linearly interpolated onto an evenly spaced grid of 256 points going around the unit circle. The log of each point was then found, giving $\hat{X}(e^{j\omega'})$. Figure 6.4-2 shows the real and imaginary parts of $\hat{X}(e^{j\omega'})$ for $\omega' = 0$ to 2π; that is, it shows the log magnitude and the phase of $X(e^{j\omega'})$.

Finally, the complex cepstrum $\hat{x}(n)$ was computed by performing an inverse FFT on $\hat{X}(e^{j\omega'})$. The resulting real-valued sequence is plotted in Figure 6.4-3. The sequence is causal, indicating that $X(e^{j\omega'})$ and $\hat{C}(j\omega)$ are minimum phase.

Since $\hat{x}(0) = .9080$, we can immediately find the scaling factor A for the best fit rational function, because $\hat{x}(0) = \log|A|$. Therefore $A = e^{.908} = 2.47935$.

The complex cepstrum $\hat{x}(n)$ was input to the DFP optimization program, which

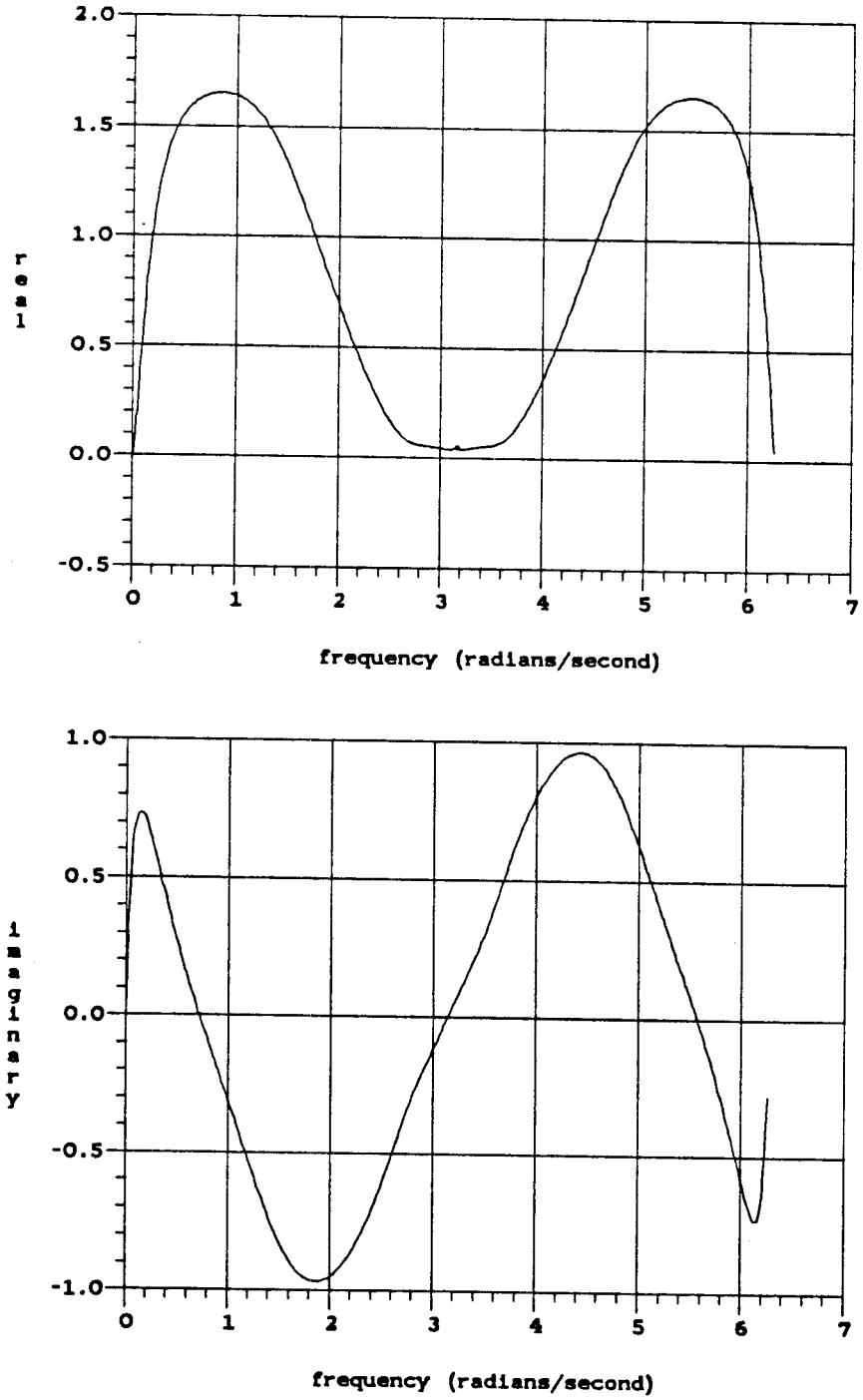

Fig. 6.4-2: Log transfer function $\hat{X}(e^{j\omega'})$ after mapping onto the unit circle. Real part is log magnitude, and imaginary part is phase in radians.

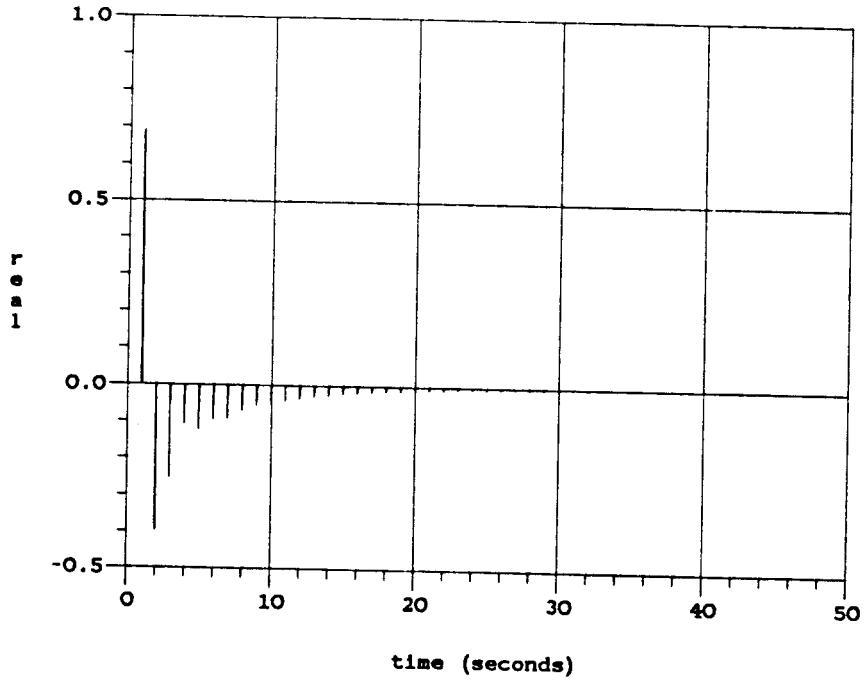

Fig. 6.4-3: The complex cepstrum $\hat{x}(n)$.

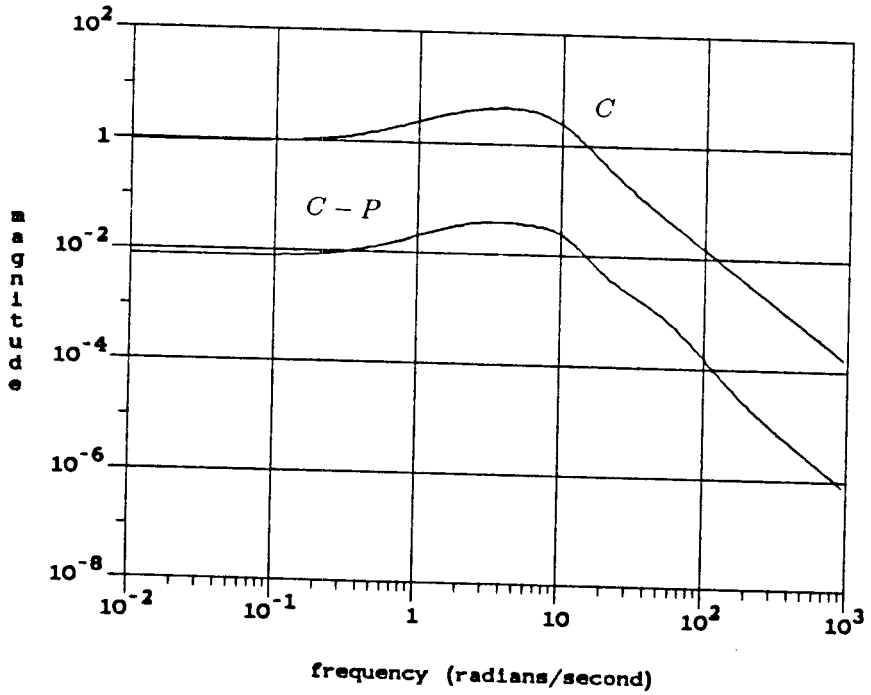

Fig. 6.4-4: Bode plots of the target C and the best fit error $C - P$. Error is approximately 1% at all frequencies.

uses the cost function and gradient of Section 6.3.1 to find a best fit rational function in z. A close fit was achieved with a third-order minimum phase transfer function with the following poles and zeros:

$$a_1 = -.366669 + j.320346 \qquad a_2 = -.366669 - j.320346 \qquad a_3 = .937329$$

$$c_1 = .113022 + j.400815 \qquad c_2 = .113022 - j.400815 \qquad c_3 = .667762$$

These poles and zeros, together with the scaling factor A, produced the rational function

$$X(z) = \frac{2.47935(z^2 + .733338z + .237068)(z - .937329)}{(z^2 - .226044z + .173427)(z - .667762)}$$

Applying the inverse Tustin transformation $z = (\alpha + s/\alpha - s)$ converts this back to the continuous time domain. The last step is to cancel the prescaling zeros with poles: multiply by $12^2/(s+12)^2$. The final result is a fifth-order rational function in s which approximates $C(j\omega)$ so closely on the $j\omega$-axis that the difference is invisible on a Bode plot. Figure 6.4-4 is a magnitude Bode plot of both $C(j\omega)$ and the approximation error $|C(j\omega) - P(j\omega)|$, where $P(s)$ is the fifth-order best fit transfer function. The error is approximately 1% at all frequencies. This corresponds to a magnitude error of 0.09 dB and a phase error of 0.6°.

The roughly constant relative error magnitude $\frac{|C-P|}{|C|}$ at all frequencies illustrates one of the advantages of matching to the log of a transfer function: the relative error, not the absolute error, is minimized at all frequencies. Minimizing the absolute error, as in H_∞-optimization, leads to a good fit at frequencies where the magnitude is large, but a poor fit where the magnitude is small.

6.5 Example: Discrete-time Compensator Synthesis

In the previous example, a nearly perfect fit was achieved for an analog transfer function target. In this next example, we will attempt to synthesize a discrete-time

rational function which makes a hybrid compensator's cone center transfer function match a target analog compensator. In this case, constraints on phase shift and stability make a perfect match impossible. We will see, however, that the complex cepstrum technique still allows for a nearly perfect match inside the loop bandwidth.

Consider the SISO feedback systems of Figures 6.5-1 and 6.5-2. The plant $G(s)$ is the same in both cases. The first system has an analog compensator $K_a(s)$. In the second system, K_a is replaced by a hybrid compensator consisting of a prefilter $F(s)$, a sampler of period $T = \pi/100 = .031416$ sec, a digital filter $K_d(z)$, and a ZOH $H_0^T(s)$. The various transfer functions are:

$$G(s) = \frac{150}{(s+1)(s+3)} \qquad F(s) = \frac{2500}{s^2+70s+2500}$$

$$K_a(s) = \frac{(s+3)^2}{(s+.4)(s+22.5)}$$

This example has been used before in [20],[23],[24]. With the exception of $K_d(z)$, this is the same system studied earlier in Section 5.1.

$K_a(s)$ is given; we wish to synthesize $K_d(z)$ so that the cone center response $\frac{1}{T}H_0^T K_d F$ approximates K_a at low frequencies. That is,

$$\frac{1}{T}H_0^T(j\omega)K_d(e^{j\omega T})F(j\omega) \approx K_a(j\omega) \qquad \text{for } -\frac{\pi}{T} < \omega < \frac{\pi}{T}$$

If this relation is satisfied, we can expect the closed-loop response and stability robustness margins to be approximately the same for both systems. Just how closely equal they are still depends on the sampling period T and the high-frequency rolloff around the loop, since these quantities affect the cone radius R. Nevertheless, if $\frac{1}{T}H_0^T K_d F \approx K_a$ inside the loop bandwidth, the closed-loop cone center responses for both systems are nearly equal.

The target function for K_d, then, is TK_a/FH_0^T. Figure 6.5-3 shows magnitude and phase plots of this target from $\omega = .01$ to $\pi/T = 100$ rad/sec. One problem is immediately apparent: the target's phase at π/T is not realizable. If $K_d(z)$

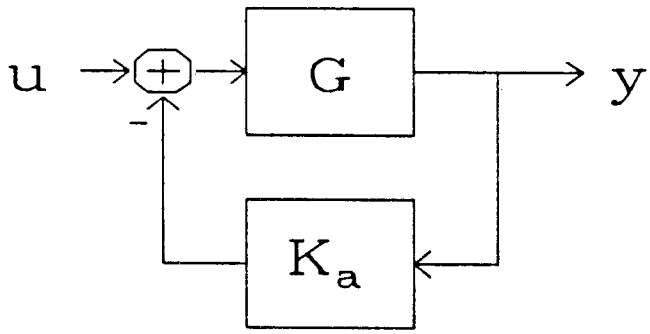

Fig. 6.5-1: An analog feedback system.

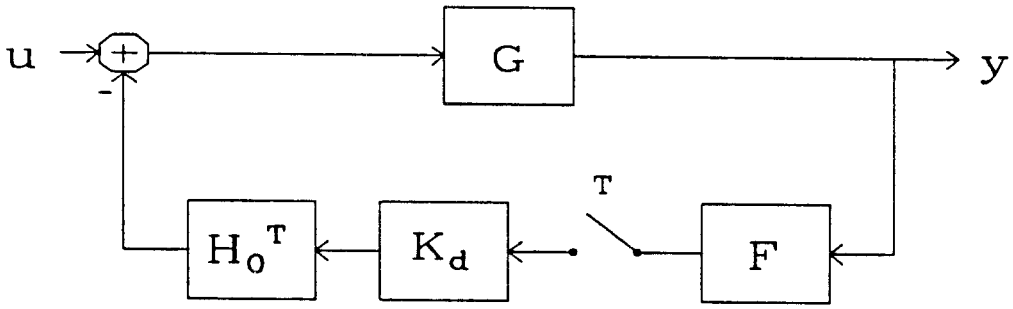

Fig. 6.5-2: The system with $K_a(s)$ replaced by a hybrid controller.

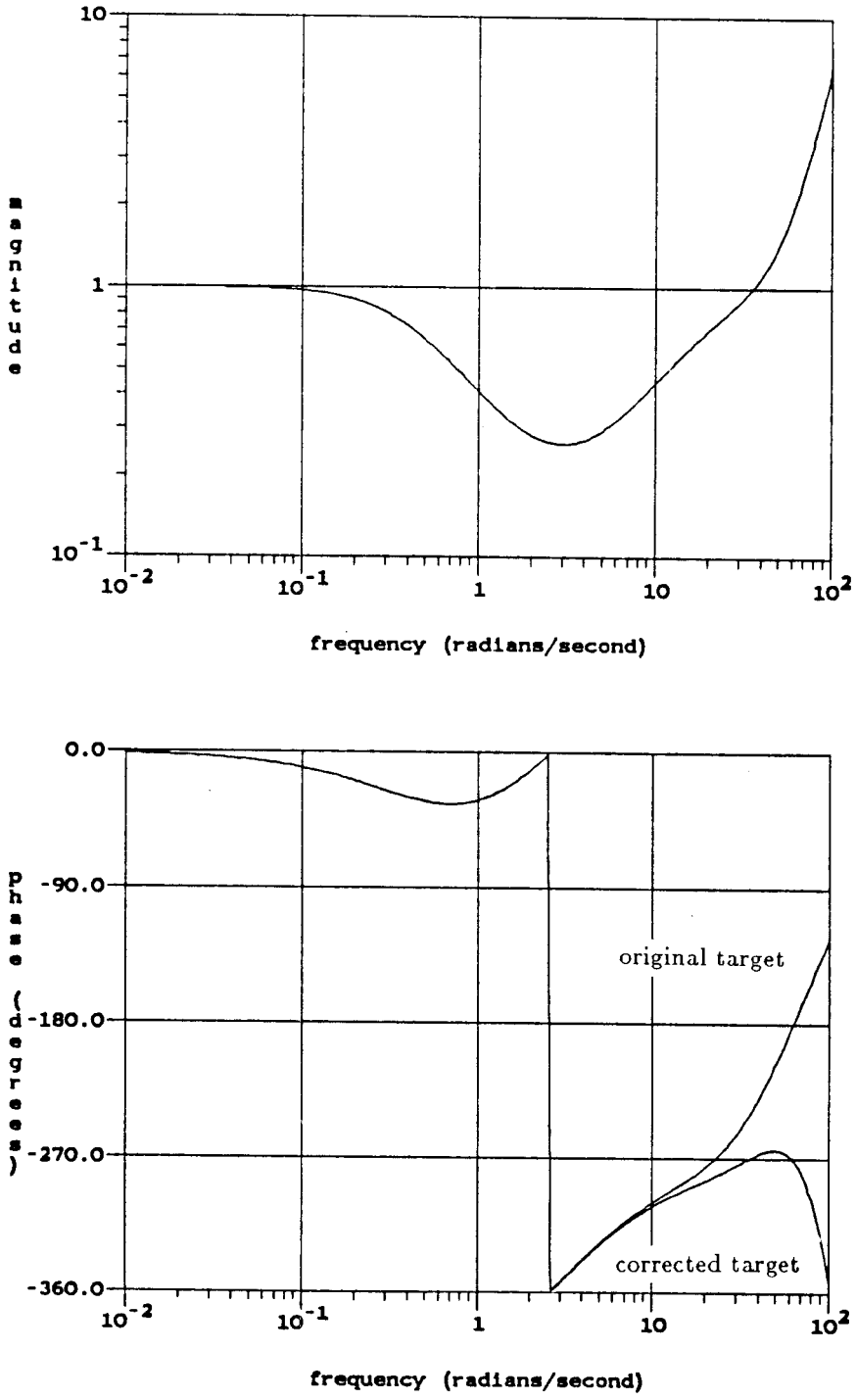

Fig. 6.5-3: Magnitude and phase of the target function, before and after phase correction.

is a rational function with real coefficients, its phase at $z = e^{j\pi} = -1$ must be either 0° or 180°. We must correct the target's phase at high frequencies. Since the phase shift is +229° = 3.997 rad at $\omega = 100$, multiply the target by $e^{-j\omega^2/\kappa}$ where $\kappa = 100^2/3.997 = 2501.9$. This has no effect on the magnitude, but it applies increasing amounts of phase lag at high frequencies, up to −229° at $\omega = 100$. The ω^2 factor causes the phase lag to grow quickly at high frequencies with little effect at low frequencies. The phase shift at loop gain crossover $\omega = 7$ is only −1.1°. Figure 6.5-3 also shows the phase of the corrected target function; the magnitude is unchanged.

Now that the corrected target is realizable, we can map it onto the unit circle, compute its complex cepstrum, and run the cepstrum-matching optimization program. The most straightforward mapping is simply to replace ω with $e^{j\omega'}$ where $\omega' = \pi/100$; this maps $\omega = 100$ to $e^{-j\pi} = -1$. However, this would crowd the entire $|\omega| < 7$ region, which we are most concerned with, into a small fraction of the unit circle. We can instead apply the Tustin transformation twice, first from z-plane to s-plane and then back again, to map $\omega = 7$ onto $z = e^{j\pi/2}$. This assigns half the unit circle to the $|\omega| < 7$ region. First replace ω' above with $\omega'' = \tan(\omega'/2)$; this maps $\omega' = \pi$ to $\omega'' = \infty$ and $\omega' = 7\pi/100$ to $\omega'' = .110$. Second, replace ω'' with $\omega''' = 2\tan^{-1}(\omega''/\alpha)$ where $\alpha = .110$; this maps $\omega'' = .110$ to $\omega''' = \pi/2$. Since $\omega'' = .110$ corresponds to $\omega = 7$, this sequence of transformations has the desired effect. Expressed in one step, the sequence is

$$\omega''' = 2\tan^{-1}\left(\frac{1}{\alpha}\tan\left(\frac{\omega\pi}{200}\right)\right)$$

It is easily verified that $\omega''' = \pi$ for $\omega = 100$ and $\omega''' = \pi/2$ for $\omega = 7$.

The importance of using the Tustin transformation is that it preserves the number of poles and zeros, their status as minimum phase or maximum phase, and the transfer function's magnitude and phase on the unit circle. The optimization pro-

gram will find pole and zero locations in the ω''' plane; we can map these back to the ω' plane easily. If z'''_0 is a pole or zero in the ω''' plane, first map it to the ω'' plane by the inverse prewarped Tustin transformation: $s''_0 = \alpha(z'''_0 - 1)/(z'''_0 + 1)$. Second, map it to the ω' plane using $z'_0 = (1 + s''_0)/(1 - s''_0)$. The final pole or zero location is given by z'_0. Expressed in one step, the complete transformation is

$$z'_0 = \frac{(1+\alpha)z'''_0 + (1-\alpha)}{(1-\alpha)z'''_0 + (1+\alpha)}$$

To summarize, first warp the corrected target, which is defined for $|\omega| \leq 100$, onto the unit circle $z = e^{j\omega'''}$. Compute the complex cepstrum for the warped target, and run the optimization program to find its poles and zeros. Finally, warp these poles and zeros back to the ω' plane.

Now we can proceed with the synthesis procedure. Sample the warped target function $X(e^{j\omega'''})$ onto an evenly-spaced grid of 256 points on the unit circle. Compute the log of each point to get $\hat{X}(e^{j\omega'''})$, then compute the inverse FFT to produce the complex cepstrum $\hat{x}(n)$. Figure 6.5-4 shows $\hat{x}(n)$, which is of course real-valued. A new problem is apparent here: $\hat{x}(n)$ is not causal. While we can still synthesize a rational function to match the anti-causal component of $\hat{x}(n)$, any such rational function will be unstable, non-causal, or both. Since the hybrid compensator must be causal and must have the same number of unstable poles as $K_a(s)$ (zero for this example), we cannot use this portion of the cepstrum.

Figure 6.5-5 shows the result of matching only the causal portion of the complex cepstrum. A third-order rational function was found to provide a very close fit; using higher-order approximations makes no significant difference. The third-order function has been transformed back into the original ω-plane and plotted together with the corrected target function. The fit is good at low frequencies, but at $\omega = 7$ a phase error of 7° is already apparent. This error is due to ignoring the non-causal portion of $\hat{x}(n)$. While fairly small, this phase error significantly reduces the

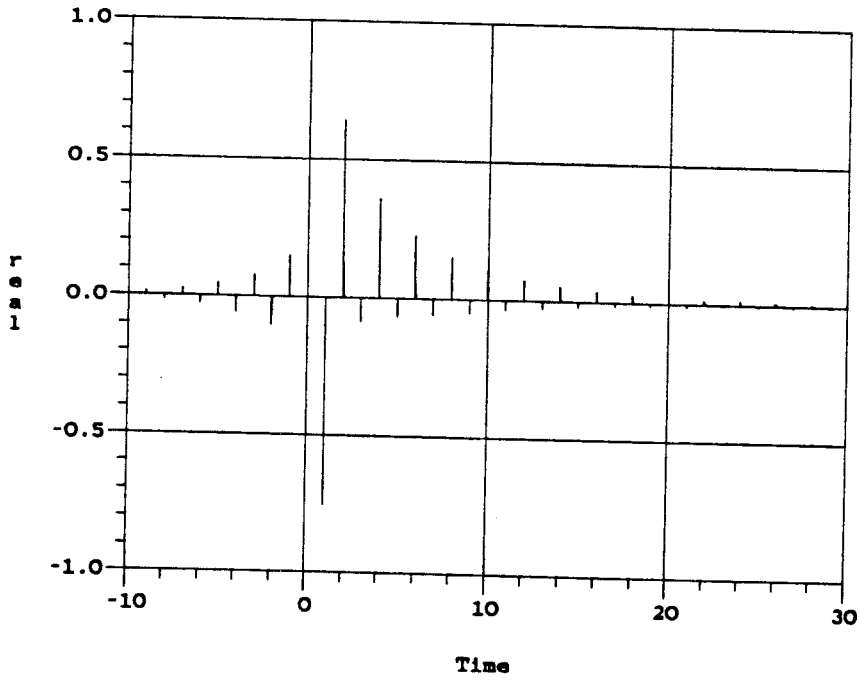

Fig. 6.5-4: Complex cepstrum $\hat{x}(n)$ of the corrected target function after mapping onto the unit circle. Noncausality implies that the target cannot be matched perfectly with a minimum phase function.

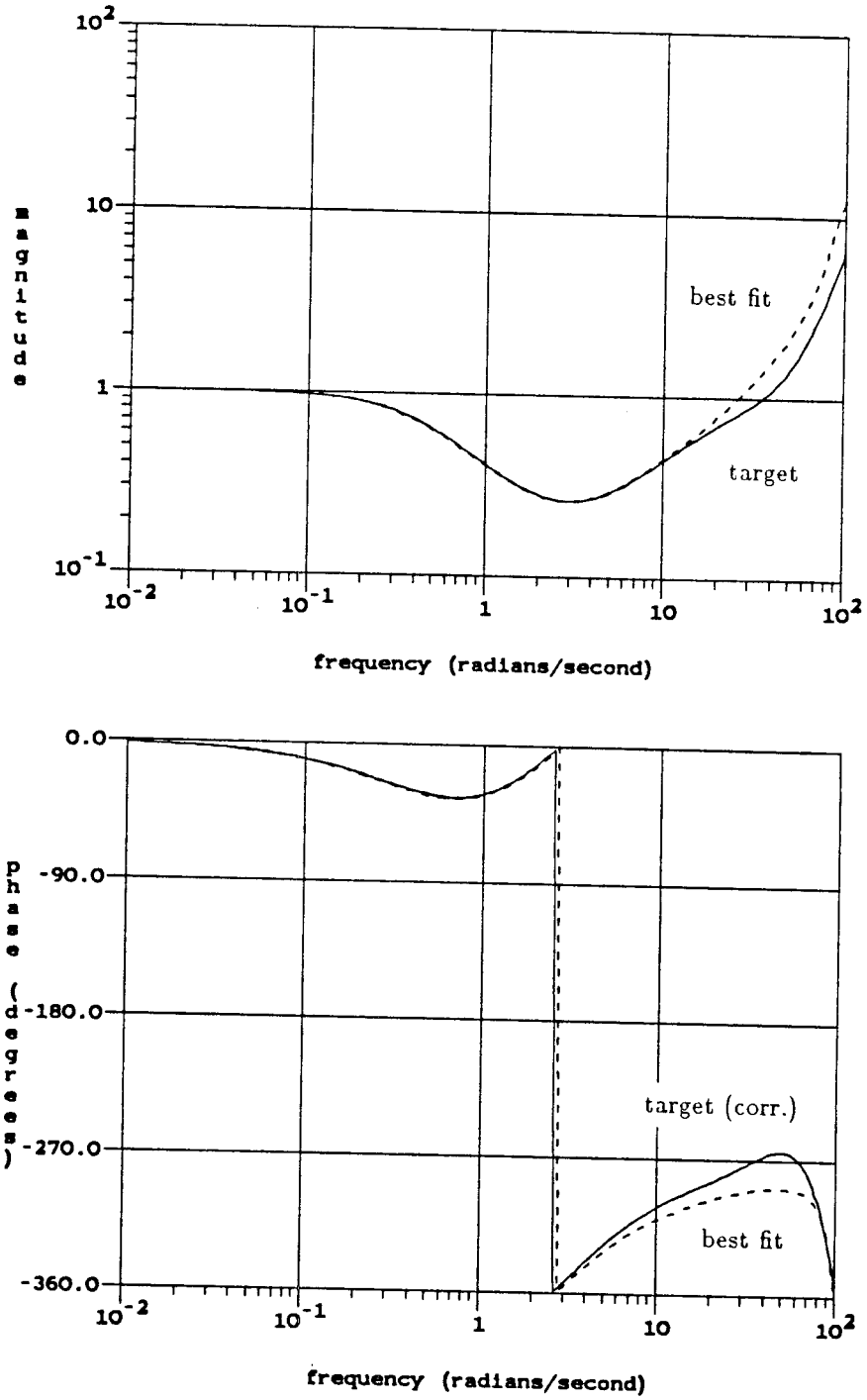

Fig. 6.5-5: Magnitude and phase of the best fit to the causal portions of $\hat{x}(n)$.

system's damping and its robustness margins. Fortunately, we can do better than this. By exploiting a basic property of the discrete Fourier transform, we can make the phase error *arbitrarily small at all frequencies*.

The key lies in realizing that once the phase of a minimum phase transfer function is given, its magnitude is predetermined. Likewise, the magnitude determines the phase. A simple modification of the complex cepstrum will produce a function with the same phase curve as the original target, but with an altered magnitude.

Any even real-valued sequence (i.e., one satisfying $\hat{x}(n) = \hat{x}(-n)$) is mapped by the FFT into another even real-valued sequence. Since the imaginary part of $\hat{X}(e^{j\omega'''})$ corresponds to phase, we can add any even real-valued sequence to the original complex cepstrum $\hat{x}(n)$ without affecting the function's phase. If we construct the sequence

$$x_2(n) = \begin{cases} -\hat{x}(-n) & n > 0 \\ 0 & n = 0 \\ -\hat{x}(n) & n < 0 \end{cases}$$

and add it to $\hat{x}(n)$, the anti-causal portion is canceled. The result is a causal sequence corresponding to a minimum phase function with exactly the same phase curve as the original target.

The optimization program was run on this modified sequence $\hat{x}(n) + x_2(n)$. Best fits were computed for second-order and third-order rational functions. The results, warped onto the original ω plane, are plotted in Figure 6.5-6, along with the corrected and uncorrected target functions. The third-order compensator achieves a nearly perfect phase match over the entire frequency range, indicating that higher-order compensators will do no better. The second-order best fit has significant high-frequency phase error, but it matches very well for $\omega < 7$. Since errors outside the loop bandwidth have little effect on the system, the second-order compensator should perform as well as the third-order one.

The magnitude plots match closely for $|\omega| < 7$, and deviate significantly from

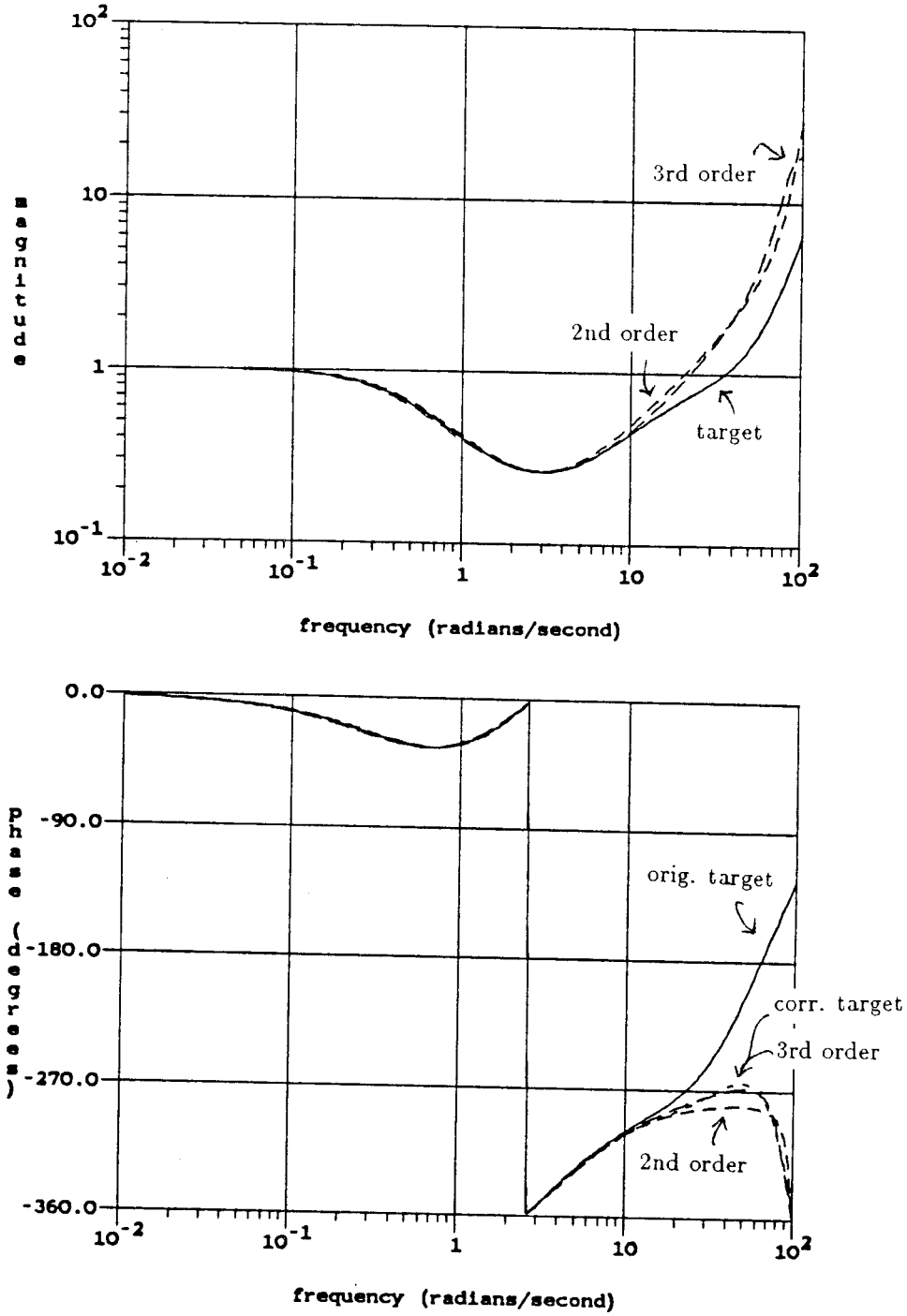

Fig. 6.5-6: Magnitude and phase of best 2nd and 3rd order fits to $\hat{x}(n)$ after correcting the target magnitude to be minimum phase. 3rd order best fit is nearly perfect.

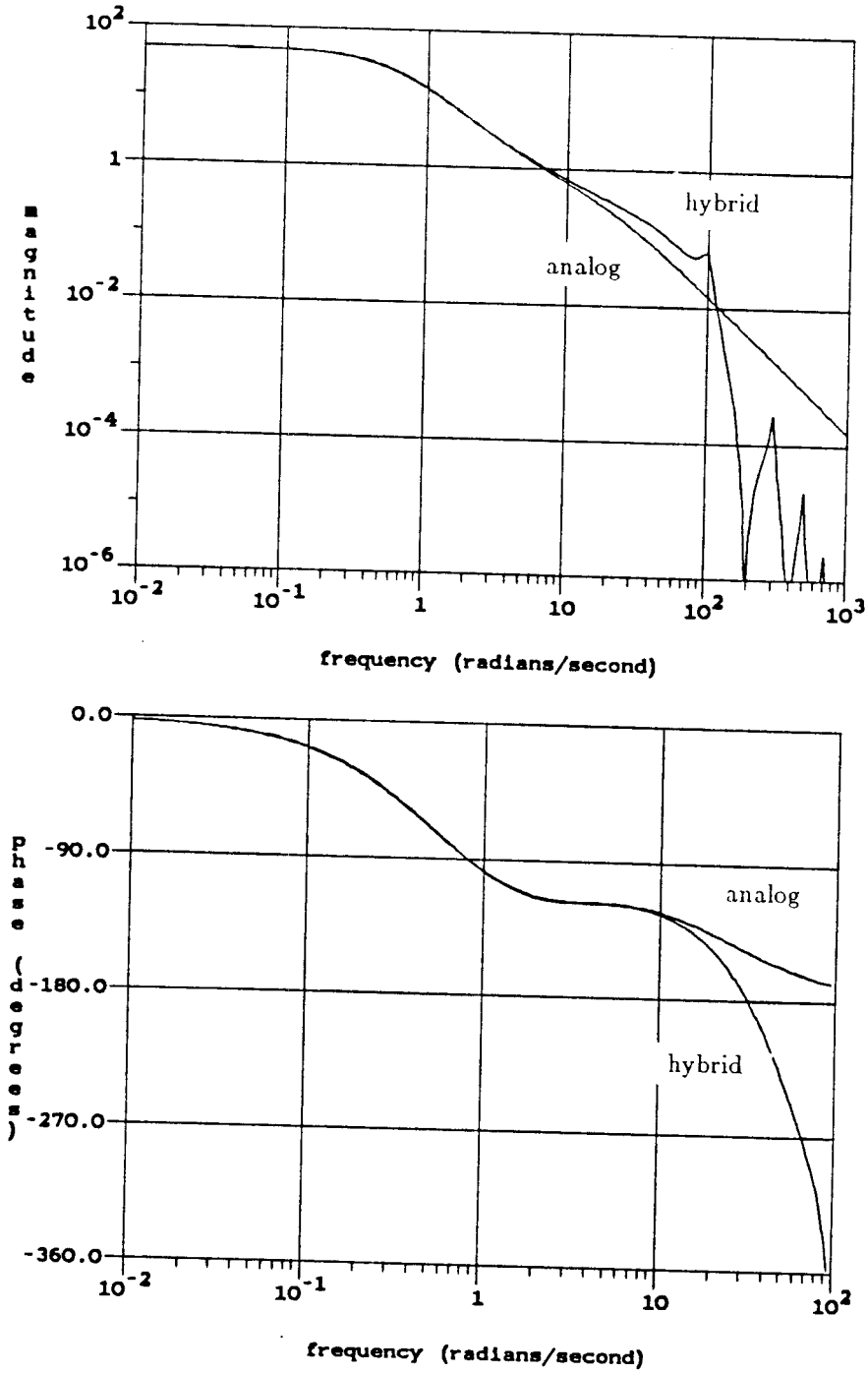

Fig. 6.5-7: Magnitude and phase of K_aG and $\frac{1}{T}H_0^TK_dFG$, the two loop gains. The match is very close inside the loop bandwidth $|\omega| < 7$.

the target at high frequencies. This is the price paid for perfect phase matching: control over the magnitude is lost. Since errors above $\omega = 7$ have no significant effect on system response, however, this is of no consequence for this example. As a practical matter, then, we have achieved an optimal match to the target with only a second-order compensator, which is given by

$$K_d(z) = \frac{2.856(1 - 1.82481z^{-1} + .8330495z^{-2})}{(1 + .806921z^{-1})(1 - .986910z^{-1})}$$

Figure 6.5-7 shows the original system's loop gain $K_a G$ and the cone center loop gain $\frac{1}{T} H_0^T K_d F G$ of the hybrid system. As expected, the magnitude and phase match very closely for $|\omega| < 7$. Figure 6.5-8 shows the step response of both systems. The hybrid system's step response closely matches that of the original analog system.

It is our tight control over the phase response of $K_d(z)$ that allows us to match the hybrid system so closely to the original analog one. By maintaining the same phase shift at loop gain crossover, we give the hybrid system approximately the same damping, overshoot, and phase margin as the target analog system. Use of the complex cepstrum provides this control by allowing us to synthesize minimum phase rational functions with any realizable phase response.

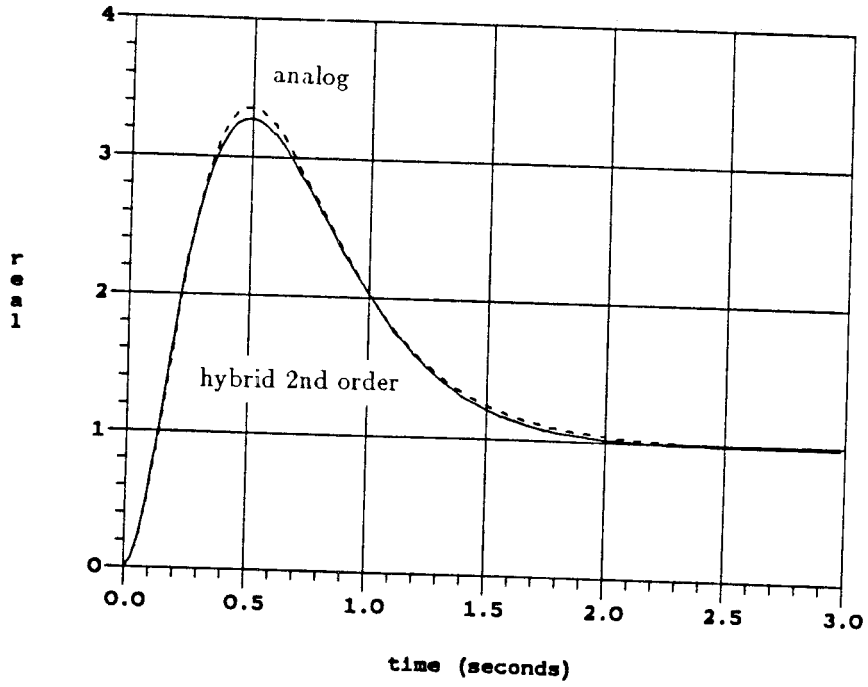

Fig. 6.5-8: Step responses of the original analog system and the sampled-data system with 2nd-order best fit digital compensator.

Chapter Seven — Conclusion

The chief contribution of this thesis is a method which greatly reduces the conservativeness of conic sector analysis for sampled-data feedback systems. Based on Theorem 3.2 from [20], which first showed how to construct a conic sector containing a hybrid operator, the new approach begins by taking advantage of the theorem's inherent flexibility in the choice of prefilter and hold. Each sampler is extracted from the system and sandwiched between fictitious LTI operators, in series with their inverses; the fictitious operators are then substituted for the prefilter and hold of the theorem. This approach was investigated earlier by Doyle and others [7]. Section 3.4 presents an optimal choice of the fictitious operators α and β for the SISO nominal stability test; the optimal choice minimizes the ∞-norm of the conic sector stability criterion $RM/\alpha\beta$.

Section 3.7 extends the SISO result to the MIMO case in a way which, though suboptimal, is nonconservative for a wide class of systems. Each sampler is extracted from the system and placed in a diagonal structure; the samplers are sandwiched between fictitious scalar operators α_i, β_i and placed in cone bounds individually. The operators α_i, β_i are chosen by applying frequency-dependent diagonal scaling to the MIMO LTI system $\tilde{M}$ to minimize the gain $\bar{\sigma}(D\tilde{M}D^{-1})$ at every frequency. This approach was inspired by the LTI Structured Singular Value (SSV), or μ-analysis, methods of Doyle [5],[6]; the only significant difference is that for LTI systems, $\bar{\sigma}(D\tilde{M}D^{-1})$ is only an upper bound for the crucial quantity μ, while in the conic sector method of this thesis, the gain $\bar{\sigma}(D\tilde{M}D^{-1})$ is an important quantity in

its own right. Software written to find the upper bound for μ may be used for this conic sector analysis without modification.

Several examples in Section 3.7.2 explore the limits of the new MIMO method, illustrating cases in which the new method is still conservative. Two of the examples are due to Doyle, and appeared previously in [4]. With one exception, it appears that conservativeness appears only when a system has poor anti-aliasing filtering, poor robustness margins, or both. To the extent that this is true, it means that the new methods will work well for well-designed systems. The exception occurs in systems with samplers at the output of a zero-order hold; this can occur in some multi-rate systems when data is sampled at one rate, held in a buffer register, then used at a different rate. Theorem 3.8 presents a modification to the basic theorem which can handle this case. Unfortunately, direct use of the modified theorem would require numerical approximation of a quadruple infinite series, a very computation-intensive enterprise. A better solution to this problem awaits further research.

It is worth noting that, by placing all samplers in an external diagonal structure, the new method makes no distinction between MIMO systems and SISO systems with multiple samplers. The total number of samplers is the only important quantity in determining the analysis structure. This means that the method makes no distinction between single-loop centralized controllers and multiple-loop decentralized systems. Furthermore, it makes no distinction between single-rate and multi-rate systems; since each sampler is initially placed in its own separate conic sector, there are no implicit assumptions on the relative timing of the samplers. Even multi-rate systems with sampling rates related by irrational numbers are included in this framework. Skewed samping effects, caused when samplers have the same rate but fire with a fixed time offset, are implicitly included. This high level of flexibility allows a large range of systems to be handled by one unified analysis method.

Section 3.9 explains how to do robustness analysis in this new framework. Plant

uncertainty is handled in the same way as in SSV analysis: all perturbations are modelled as normalized Δ-blocks, extracted from the system and arranged in a block-diagonal structure. This approach to structured uncertainty has great power and is much less conservative than older approaches to MIMO model uncertainty; it also merges naturally with the new conic sector methods of this thesis. All cone-bounded samplers and LTI Δ-blocks are arranged into a single block-diagonal operator and placed in a single collective conic sector; then the standard conic sector stability criterion can be applied.

The basic system tests are the same as in SSV analysis: nominal stability, robust stability, nominal performance, and robust performance. The performance tests amount to finding a conic sector containing the closed-loop system, either the nominal system or the set of all allowed perturbed systems. Chapter 4 explains how to use such a closed-loop cone to test specifications on RMS response to sine wave and white noise inputs. The closed-loop cone center may be used to estimate the closed loop step response; an example of this is given in Chapter 5. Unfortunately, no error bounds, in the form of an envelope containing the step response, can be given. Chapter 4 also describes how to compute lower bounds for the gain margin, phase margin, and gain-phase margins in the conic sector framework. Sections 4.1 and 4.2 discuss algorithms for computing the cone radius and for performing diagonal scaling.

In Chapter 5, three examples were presented: a SISO system with one sampler, a MIMO single rate system, and a multi-rate system. The first example was examined in detail, to illustrate most of the analysis tests of Chapters 3 and 4. For the second and third examples, the nominal and robust stability tests were made. The first two examples provided extremely nonconservative results. The conic sector based gain-phase margin of the SISO system was only 0.4% smaller than the margin computed by standard z-transform techniques. For the MIMO system, the conic sector

based margin for constant multiplicative perturbations was only 0.7% smaller than the z-transform based margin. In return for this small degree of conservativeness, the conic sector methods gave a frequency-dependent bound on allowable plant perturbations, which is much more useful than the constant bound from z-transform methods.

The multi-rate example gave more conservative results, with a conic sector based gain-phase margin 30% smaller than the margin derived from Kranc operator analysis. Again, however, the conic sector based margin is based on the realistic assumption of frequency-dependent plant perturbations, not on the assumed constant perturbations of the Kranc method and of standard z-transform methods. Therefore the frequency-dependent conic sector based robustness margin has greater practical use.

The second chief contribution of this thesis is presented in Chapter 6; it is a synthesis method for SISO hybrid compensators based on the complex cepstrum. The philosophy is the same as that of previous work by Thompson, Sánchez Peña, and Wong [24]: to synthesize a digital filter $K_d(z)$ so that the hybrid compensator's cone center $\frac{1}{T}H(j\omega)K_d(e^{j\omega T})F(j\omega)$ closely approximates a target analog compensator $K_a(j\omega)$ for ω inside the loop bandwidth. This approach compensates for the phase lag introduced by the hold H and prefilter F; this phase lag has been ignored in many conventional discretization techniques and often leads either to inferior compensators or to excessively high sampling rates.

The target function TK_a/HF is computed and its phase near $\omega = \pi/T$ is adjusted to make it realizable. Then the target function is mapped onto the unit circle and its complex cepstrum is computed via the complex logarithm and the FFT. An appropriate cost function is defined and a nonlinear optimization problem is solved to find the best fit. Since both the cost function and its gradient are simple functions of the estimated pole and zero locations, gradient-based iterative methods

such as the Davidon-Fletcher-Powell (DFP) algorithm can solve the problem with relative efficiency.

In an example in Chapter 6, the complex cepstrum method provides an excellent fit to a second-order compensator with a second-order discrete-time compensator. The significant in-band phase lag of the prefilter and hold is completely cancelled out without resorting to a higher order compensator. This excellent control over phase shift is a consequence of a basic property of Fourier transforms, which allows us to synthesize a minimum-phase function with any desired phase response.

In a second example, another application of the complex cepstrum synthesis method is shown: fitting rational functions of given order to a set of magnitude and phase points of an analog (or discrete) transfer function. The data points can be actual measurements or computed values, as in the example. A fifth-order rational function in s is found which closely approximates the closed loop cone center of the example system from Section 5.1. Since the cone center contains $e^{j\omega T}$ terms, it is actually infinite-dimensional; the fifth-order best fit is seen to approximate it extremely well at all frequencies.

This method is equally useful for matching both analog and discrete transfer functions, since one can be mapped into the other by the bilinear, or prewarped Tustin, transformation.

The third contribution of this thesis is a state-space realization for Kranc operators, used in multi-rate system analysis. This material was presented in Section 2.8 at the end of the chapter on mathematical background, since it has appeared earlier in [22], and since it merges naturally with the standard z-transform methods of Section 2.7. Kranc switch decomposition consists of replacing a sampler of period T/n with n samplers of period T, each in series with appropriate time delays and advances. In this way each subsystem of a multi-rate feedback system is replaced by a MIMO single-rate system, for which a transfer function is defined. Then stan-

dard MIMO z-transform methods can be used to check stability. When done in the z-transform domain, however, this process can lead to large matrices of high-order rational functions in z, which are difficult to work with. In the state-space domain, however, such MIMO Kranc operators are easy to compute in the form of a state-space quadruple. Furthermore, the number of states of a subsystem's Kranc operator is the same as that of the original subsystem, if zero-order holds are used. This preservation of the number of states is the payoff for state-space Kranc analysis; it keeps the problem complexity from growing out of control as the number of inputs and outputs is increased.

Closed-loop stability is checked by first interconnecting the state-space Kranc operator representations, a simple process; then by computing the maximum eigenvalue of the closed loop state transition matrix. Time domain input-output simulation is trivial with a state-space quadruple for the closed loop system's Kranc operator. Fictitious high-speed samplers can even be placed at the system's analog outputs to check intersample ripple behavior.

In addition to the state-space realization for Kranc operators, Section 2.8 also presented Thompson's methods [21] for using Kranc operators to find robustness margins for multi-rate systems. These results were later used in Section 5.3 to test the conservativeness of the conic sector methods for a multi-rate example. Section 5.3 also gave two explicit examples of the calculation of state-space quadruples for Kranc operators.

7.1 Suggestions for Future Research

1. MIMO complex cepstrum synthesis. The complex cepstrum synthesis method of Chapter 6 can be used only to find best fits to SISO transfer functions, since the complex cepstrum is parameterized in terms of SISO poles and zeros.

An equivalent complex cepstrum for a matrix-valued transfer function can be defined and computed easily enough, by simply finding the complex cepstrum of each matrix element and arranging the results in a new matrix. To use this for synthesis, however, would require a parameterization in terms of some underlying MIMO structure, either a state-space quadruple or appropriately defined MIMO poles and zeros.

If such a parameterization is found and is usable, it could have great utility not only in MIMO digital compensator design but also in MIMO model reduction, which is now something of a black art. Modern analog synthesis techniques, such as H^∞ methods and μ-synthesis, tend to produce compensators of very high order which must then be approximated by low-order functions.

2. Synchronized sampling. The approach of this thesis in placing every sampler in its own individual conic sector allows us to handle skewed and multirate sampling schemes with no additional effort. However, for systems which have known, unchanging time offsets between their various samplers (most systems do) this can potentially lead to conservativeness, since useful information is being discarded. This situation might be improved by using non-diagonal scaling factors, which are equivalent to crossfeeds between channels which are known to be synchronized. The inverse of such a scaling factor, placed after the samplers, removes the effects of the crossfeed if the scaling factor is constant or periodic in ω. The additional freedom of allowing off-diagonal elements in D could further reduce the gain $\inf_D \bar{\sigma}(D\tilde{M}D^{-1})$ and lead to a smaller cone radius.

3. Optimal choice of A, B in MIMO case. The method of Section 3.7, which finds $\hat{\mu} = \inf_D \bar{\sigma}(D\tilde{M}D^{-1})$ and sets $A = \hat{\mu}^{1/2} = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ and $B = \hat{\mu}^{1/2} = \mathrm{diag}(\beta_1, \ldots, \beta_n)$ produces a guaranteed upper bound $\bar{\sigma}(R)$ for the conic sector stability criterion $\bar{\sigma}(R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2})$. However, it does not guarantee to minimize $\|R^{1/2}A^{-1}\tilde{M}B^{-1}R^{1/2}\|_\infty$, so it is a suboptimal solution. A systematic

way to find the optimal choice for A and B would further reduce this method's conservativeness.

4. Less conservative sector methods for nonlinear systems. Conic sector theory, and Safonov's generalized sector theory of which it forms a part, are formulated to allow nonlinear system analysis. Unfortunately, conic sectors generally produce very conservative bounds for many typical nonlinearities, including amplifier saturation and the quantization that occurs in A/D converters. Perhaps the more general sector theory of [16] can be employed to find less conservative bounds for these commonly encountered nonlinearities; the results could be useful in determining, for instance, the required number of bits for an A/D converter or a digital filter.

References

[1] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking," *Proc. Symp. Time Series Analysis*, M. Rosenblatt, Ed., New York, John Wiley and Sons, New York, 1963, pp. 209-243.

[2] W. H. Boykin and B. D. Frazier, "Multirate Sampled Data Systems Analysis Via Vector Operators," *IEEE Trans. Auto. Control*, Vol. AC-20, pp. 548-551, Aug. 1975.

[3] J. C. Doyle and G. Stein, "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," *IEEE Trans. Auto. Control*, Vol. AC-26, pp. 4-16, Feb. 1981.

[4] J. C. Doyle, "More Conic Sector Analysis of Hybrid Control Systems," Honeywell Interoffice Correspondence, Jan. 11, 1983.

[5] J. C. Doyle, "Analysis of feedback systems with structured uncertainty," *IEE Proceedings, Part D*, Vol. 129, No. 6, Nov. 1982.

[6] J. C. Doyle, "Lecture Notes," 1984 ONR/Honeywell Workshop on Advances in Multivariable Control, October 8-10, 1984, Minneapolis, MN.

[7] J. E. Wall, J. C. Doyle, G. L. Hartmann, N. A. Lehtomaki, G. Stein, "Performance and Robustness Aspects of Digital Control Systems," internal publication, Honeywell Systems and Research Center, Minneapolis, MN.

[8] M. K. H. Fan and A. L. Tits, "Characterization and Efficient Computation of the Structured Singular Value," *IEEE Trans. Auto. Control*, Vol. AC-31, No. 8,

pp. 734-743, August 1986.

[9] R. Fletcher and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," *Computer J.*, Vol. 6, No. 2, pp. 163-168, 1963.

[10] R. E. Kalman and J. Bertram, "A Unified Approach to the Theory of Sampling Systems," *J. Franklin Inst.*, Vol. 267, pp. 405-436, May 1959.

[11] A. Kostovetsky, "Some Investigations of Hybrid Systems," S.M. Thesis, Dept. of Mechanical Engineering, MIT, May 1970, Appendix to Lab. for Info. and Decision Systems Report LIDS-FR-960.

[12] G. M. Kranc, "Input-Output Analysis of Multirate Feedback Systems," *IRE Trans. Auto. Control*, Vol. AC-3, pp. 21-28, Nov. 1957.

[13] E. E. Osborne, "On Preconditioning of Matrices," *J. Assoc. Comp. Mach.*, Vol. 7, pp. 338-345, 1960.

[14] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[15] M. G. Safonov, "Propagation of Conic Model Uncertainty in Hierarchical Systems," *IEEE Trans. Auto. Control*, Vol. AC-28, pp. 701-709, June 1983.

[16] M. G. Safonov, *Stability and Robustness of Multivariable Feedback Systems*, The MIT Press, Cambridge, Massachusetts, 1980.

[17] M. G. Safonov and M. Athans, "A Multiloop Generalization of the Circle Criterion for Stability Margin Analysis," *IEEE Trans. Auto. Control*, Vol. AC-26, pp. 415-422, April 1981.

[18] M. G. Safonov, "Optimal Diagonal Scaling for Infinity-Norm Optimization," *Systems and Control Letters*, Vol. 7, No. 4, pp. 257-260, July 1986.

[19] G. Stein, M. Athans, and P. M. Thompson, *Hybrid Operator Models for Digitally Implemented Control Systems*, Final Report for NASA Langley Research Center, NASA Grant No. NAG1-2, MIT Lab. for Info. and Decision Systems Report LIDS-FR-1206, April 1982.

[20] P. M. Thompson, *Conic Sector Analysis of Hybrid Control Systems*, Ph. D. Thesis, Dept. of Electrical Engineering and Computer Science, MIT, Sept. 1982.

[21] P. M. Thompson, "Gain and Phase Margins of Multi-Rate Sampled Data Feedback Systems," *Intl. J. Control*, Vol. 44, No. 3, pp. 833-846, 1986.

[22] P. M. Thompson and R. L. Dailey, "Kranc Vector Switch Decomposition in State Space," *J. Guidance, Control, and Dynamics*, submitted for publication.

[23] P. M. Thompson, R. L. Dailey, and J. C. Doyle, "New Conic Sectors for Sampled-Data Feedback Systems," *Systems and Control Letters*, Vol. 7, No. 5, pp. 395-404, Sept. 1986.

[24] P. M. Thompson, R. Sánchez Peña, and Y.-F. Wong, "Design of Sampled-Data Compensators Based on the Use of Conic Sectors," *Proceedings of the 1986 American Control Conference*, Seattle, WA, June 18-20, 1986, pp. 1462-1464.

[25] V. M. Walton, "State Space Stability Analysis of Multirate Multiloop Sampled Data Systems," *AAS/AIAA Astrodynamics Specialist Conference*, Lake Tahoe, Nevada, August 3-5, 1981.

[26] R. F. Whitbeck, *Elements of Simulation Error Analysis*, Systems Technology, Inc., Final Report for Air Force Flight Dynamics Laboratory, Report AFWAL-TR-82-3022, June 1982.

[27] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems, Part I," *IEEE Trans. Auto. Control*, Vol. AC-11, pp. 228-238, April 1966.

[28] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems, Part II," *IEEE Trans. Auto. Control*, Vol. AC-11, pp. 465-476, July 1966.

[29] P. M. Thompson, G. Stein, M. Athans, "Conic Sectors for Sampled-Data Feedback Systems," *Systems and Control Letters*, Vol. 3, pp. 77-82, July 1983.