

Optimization Methods in Analysis and Design of Linearized Systems*

April 22, 2005

*©A. Megretski, 2005. All rights reserved.

Contents

1	Introduction	7
1.1	Motivating examples	7
1.1.1	Inverted pendulum with delayed control	7
1.1.2	Efficient MIMO stabilization	11
1.2	Structure of the lectures	12
1.2.1	Systems and their approximation	12
1.2.2	Explicit optimization of linear systems	13
1.2.3	Convex optimization and robustness analysis	13
2	System models	14
2.1	General systems	14
2.1.1	Signals	14
2.1.2	Behavioral models	16
2.1.3	L2 power gain	17
2.1.4	Outer approximations of systems	18
2.1.5	Small gain theorem	20
2.2	Transfer matrix models of LTI systems	21
2.2.1	Laplace transform and continuous time Fourier transform	21
2.2.2	Transfer matrix models of continuous time systems	23
2.2.3	H-Infinity norm	25
2.2.4	L2 gain of transfer matrix models	26
2.2.5	Outer approximations of transfer matrix models	28
2.3	State space models	30
2.3.1	Definition of state space LTI models	30
2.3.2	Controllability and observability	31
2.3.3	L2 gain of state space models	34
3	Classical optimization schemes	36
3.1	Feedback stabilization setup	36
3.1.1	Plant, controller, closed loop system	36
3.1.2	On the assumption $D_{22} = 0$	37
3.1.3	Transfer matrices of feedback interconnection	39
3.1.4	Zero/pole cancellation: an example	40
3.2	Norm minimization schemes and singularity	41
3.2.1	Well posedness of norm minimization problems	42
3.2.2	Examples involving different types of singularity	42

3.3	H-Infinity optimization	44
3.3.1	H-Infinity norm and its calculation	44
3.3.2	Entropy minimization setup	45
3.3.3	Example: non-unique H-Infinity optimal controller	45
3.3.4	Example: L2 gain optimization via small gain theorem	46
3.4	H2 optimization setup	48
3.4.1	H2 norm	49
3.4.2	H2 norms and response to Gaussian white noise	50
3.4.3	H2 optimization and its singularity	51
3.5	Hankel optimal model reduction setup	53
3.5.1	The Hankel norm	53
3.5.2	Hankel optimal model reduction	53
4	Q-parameterization and the “waterbed effect”	56
4.1	Q-Parameterization	56
4.1.1	The setup	56
4.1.2	The Q-parameterization theorem	57
4.1.3	Derivation of Youla parameterization	58
4.1.4	Open loop zeros	59
4.1.5	Time-Domain Q-Parameterization	60
4.2	The waterbed effect	61
4.2.1	The Parseval Identity	61
4.2.2	Integral Identities for Stable Transfer Functions	62
4.2.3	Example: Poisson formula and closed loop bandwidth	63
5	Kalman-Yakubovich-Popov Lemma	66
5.1	Motivating examples	66
5.1.1	Completion of squares in optimal program control	66
5.1.2	Quadratic storage functions as L2 power gain certificates	69
5.1.3	Differential games	70
5.2	Systems constructions related to the KYP Lemma	71
5.2.1	Riccati equations and inequalities	72
5.2.2	Frequency domain inequalities and spectral factorization	73
5.2.3	Linear matrix inequalities	74
5.2.4	Riccati equations and stabilizing solutions	76
5.2.5	Stable invariant subspaces of Hamiltonian matrices	77
5.2.6	Abstract H2 optimization and completion of squares	79
5.3	Main statements of the KYP lemma	80

5.3.1	Checking generalized positivity	81
5.3.2	Checking positivity	82
5.3.3	A lemma on Hamiltonian matrices	83
5.3.4	Stabilizing solutions of ARE with stabilizable $\theta \geq 0$	84
5.3.5	Stabilizing solutions of the dual ARE	85
6	H2 optimization	87
6.1	Derivation of H2 optimal controller	87
6.1.1	Background: impulse responses	87
6.1.2	Closed loop impulse responses	88
6.1.3	Reduction to abstract H2 optimization	89
6.1.4	Optimal H2 controller	90
6.2	Properties of H2 optimal controllers	91
6.2.1	Closed Loop Poles	91
6.2.2	Closed Loop Stability Robustness	93
7	H Infinity optimization	94
7.1	Problem Formulation and Algorithm Objectives	94
7.1.1	Suboptimal Feedback Design	94
7.1.2	Why Suboptimal Controllers?	95
7.1.3	What is done by the software	96
7.2	Background Results	97
7.2.1	A special case of the KYP Lemma	98
7.2.2	The Parrot's Lemma and Its Generalizations	98
7.3	H-Infinity Optimization for a Simplified Setup	100
7.3.1	The Simplified Setup Solution	100
7.3.2	Proof of Theorem 7.4	102
8	Robustness analysis with quadratic constraints	104
8.1	General principles of IQC analysis	104
8.1.1	Definition of IQC	104
8.1.2	IQC for the pure integrator	104
8.1.3	System analysis via IQC and convex optimization	105
8.1.4	Example: small gain theorem via IQC manipulations	106
8.1.5	Example: L2 gain bound with multiple uncertain subsystems	107
8.1.6	Dynamic extensions	108
8.2	Quadratic constraints for LTI models	109
8.2.1	Zero Exclusion Principle	109

8.2.2	Structured Singular Values	110
8.2.3	Examples	111
8.2.4	A “Small μ Theorem”	111
8.3	Computation of robustness margins	112
8.3.1	Lower Bounds of μ	112
8.3.2	Quadratic Constraints	113
8.3.3	Elementary Uncertainty	114
8.3.4	An Upper Bound for μ	114
8.4	Numerical calculations and examples	115
8.4.1	Example: a servo with friction and uncertain delay	115
8.4.2	Example with cubic nonlinearity and delay	117
9	Model order reduction	121
9.1	Objectives and challenges of model reduction	121
9.1.1	Motivating example: the heat equation	121
9.1.2	Motivating example: system identification	122
9.1.3	Simplification of general system models	122
9.1.4	Example: reduction of matrix-vector products	123
9.1.5	Challenges of model reduction	125
9.2	Model reduction by projection	126
9.2.1	Projection of finite order state space LTI models	126
9.2.2	An alternative interpretation of projection MOR	128
9.2.3	Projections for other model types	129
9.2.4	Preservation of transfer matrix moments	131
9.2.5	Stability preservation	134
9.2.6	L2 gain and passivity preservation	137
9.3	Balanced Truncation	139
9.3.1	Motivation: removal of (almost) uncontrollable/unobservable modes	139
9.3.2	Observability measures	140
9.3.3	Controllability measures	143
9.3.4	Joint controllability and observability measures	144
9.3.5	Classical balanced truncation	146
9.3.6	Approximate Gramians	147
9.3.7	Lower bounds for model reduction error	149
9.3.8	Upper bonds for balanced truncation errors	149
9.4	Hankel Optimal Model Reduction	153
9.4.1	Hankel operators	153
9.4.2	Hankel matrices	154

9.4.3	Singular values of a Hankel operator	155
9.4.4	The AAK theorem	157
9.4.5	AAK theorem proof: explicit formulae and certificates	161
9.4.6	KYP lemma for L-Infinity norm approximation	163
9.4.7	Hankel optimal reduced models via Parrot's Theorem	164
10	Convex optimization	167
10.1	Basic theory of convex analysis	167
10.1.1	Convexity	167
10.1.2	Dual representations of convex sets and functions	168
10.1.3	Example: a convex function	169
10.1.4	Second derivative and convexity	170
10.1.5	Convexity-preserving operations	170
10.1.6	The Hahn-Banach Theorem	172
10.1.7	Duality gap for linear programs	172

1 Introduction

The main objective of these lectures is to provide a complete but compact introduction to *modern* linear control systems, which, in contrast with the *classical* approach, seeks to utilize rigorous mathematical methods and computer-aided optimization in design and analysis of feedback systems. The presentation will explain basic principles, mathematical results, and numerical implementation strategies related to the use of optimization in design and analysis of dynamical systems which are either linear or can be dependably approximated by linear systems. The basic principles include use of L2 gains or Integral Quadratic Constraints for quantifying quality of linear approximations and feedback system performance, use of generalized small gain conditions and convex optimization in analysis of robustness to modeling errors and uncertain parameters, and use of Schur decomposition for explicit optimization of linear feedback and reduced models of linear systems. Among the remarkable mathematical results associated with the theory are Kalman-Yakubovich-Popov lemma describing the relation between frequency domain inequalities and stabilizing solutions of Riccati equations, Adamyan-Arov-Krein theorem on Hankel optimal model reduction, and the state space solution of H-Infinity optimization problem. Practical implementation of these basic concepts and mathematical theory rely on numerical linear algebra and convex optimization engines: linear equation solvers, Schur decomposition, and LMI optimization (semidefinite programming).

1.1 Motivating examples

Typically, the need to use modern linear control arises when working with models which are “complex” (no good second order approximation) or MIMO (multiple input, multiple output), or when optimization of performance is a concern. Application of modern linear control to a set of system optimization questions is sketched here, though complete solutions will have to wait until the essential techniques are developed in the coming sections.

1.1.1 Inverted pendulum with delayed control

Consider the standard equations for inverted pendulum with no friction (Figure 1.1):

$$\ddot{\theta}(t) = \omega_0^2 \sin(\theta(t)) + v(t - T), \quad (1.1)$$

where $\theta(t)$ is the angular position, $\omega_0^2 = g/L > 0$ is the ratio of free fall acceleration and pendulum’s length, $v(t)$ is the control torque to mass ratio, and $T \geq 0$ is the control loop delay.

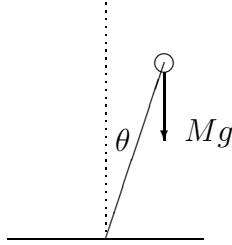


Figure 1.1: Inverted pendulum

Assuming that a and T are known, a basic task of linear control could be to stabilize *locally* the upright position of the pendulum by finding a finite order LTI (linear time invariant) *feedback law*,

$$v(t) = C_f x_f(t) + D_f \theta(t), \quad (1.2)$$

$$\dot{x}_f(t) = A_f x_f(t) + B_f \theta(t), \quad (1.3)$$

where A_f, B_f, C_f, D_f are constant real coefficient matrices, and $x_f = x_f(t)$ is the *state* of the feedback system. The term “local” refers to using an intentionally weakened stabilization requirement: all solutions $(\theta(t), x_f(t))$ of (1.1)-(1.3) with initial conditions $\theta(0), \dot{\theta}(0), x_f(0)$ sufficiently close to zero should converge to zero as $t \rightarrow \infty$.

In this example, the original model is “complex”, in the sense that it involves a non-linear component $\theta(t) \mapsto \sin(\theta(t))$ and an infinite order LTI component $v(t) \mapsto v(t - T)$. Application of modern control principles to the stabilization problem would begin with finding good finite order LTI approximation of (1.1), complete with error bounds, expressed in terms of *L2 gains*.

Memoryless nonlinear transformations, such as $\theta(t) \mapsto \sin(\theta(t))$, cannot be approximated by the linear ones with arbitrary accuracy. As a compromise, higher quality of approximation in a certain region of values of the input is achieved at the expense of lowering the quality in the complementing region. If the region of admissible values of $\theta(t)$ is $[-\theta_0, \theta_0]$, where $\theta_0 \in (0, \pi)$, one can use $\sin(\theta(t)) \approx k\theta(t)$, where

$$k = \frac{1}{2} + \frac{\sin(\theta_0)}{2\theta_0},$$

defines the center line $\{(t, k\theta) : \theta \in \mathbf{R}\}$ of the (non-convex) cone spanned by the points $(\theta, \sin(\theta))$ with $|\theta| \leq \theta_0$. This leads to a finite order LTI representation

$$\sin(\theta(t)) = k\theta(t) + d_1^{-1}w_1(t), \quad (1.4)$$

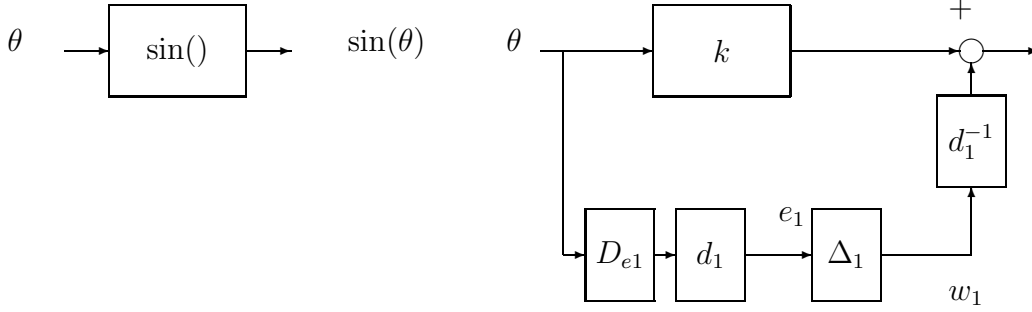


Figure 1.2: Linearization of $\sin(\cdot)$

visualized by the block diagram of Figure 1.2, where w_1 is a scaled *modeling error signal*, Δ_1 is the *modeling error system*, mapping e_1 to w_1 ,

$$w_1 = d_1(\sin(\theta) - k\theta), \quad (1.5)$$

$$e_1(t) = d_1 D_{e1} \theta(t), \quad (1.6)$$

$$D_{e1} = 0.5(1 - \sin(\theta_0)/\theta_0).$$

A simple calculation shows that

$$\int_0^T |w_1(t)|^2 dt \leq \int_0^T |e_1(t)|^2 dt, \quad (1.7)$$

provided that $|\theta(t)| \leq \theta_0$ for all t . Condition (1.7) serves as an *L2 gain bound* for the modeling error system Δ_1 , quantifying the approximation error. The theory of approximating nonlinear systems by the linear ones, and using conditional energy gain bounds in stability and performance analysis is a component of modern control.

For the delay transformation $v(t) \mapsto v(t - T)$, one can use representation of the form

$$v(t - T) = \hat{v}(t) + d_2^{-1} w_2(t), \quad (1.8)$$

$$\hat{v}(t) = C_d x_d(t) + D_d v(t), \quad (1.9)$$

$$\dot{x}_d(t) = A_d x_d(t) + B_d v(t), \quad (1.10)$$

where A_d, B_d, C_d, D_d are coefficient matrices of an LTI state space model approximating the delay, $w_2(t)$ is scaled delay modeling error, and $d_2 > 0$ is a scaling parameter. A pure delay cannot be approximated by a finite order LTI system with arbitrary accuracy,

because the error is always large for high frequencies. As a compromise, the approximation error is quantified by establishing an energy bound of the form

$$\int_0^T |w_2(t)|^2 dt \leq \int_0^T |e_2(t)|^2 dt, \quad (1.11)$$

where

$$e_2(t) = d_2(C_{e2}x_d(t) + D_{e2}v(t)) \quad (1.12)$$

is an auxiliary output of (1.10), defining e_2 as the result of applying a low-pass filter to v . For example, when $T > 0$ is small enough, which makes the delay easier to approximate, the coefficients in (1.9),(1.10),(1.12) can be defined by

$$D_d + C_d(sI - A_d)^{-1}B_d = \frac{1 - Ts/2}{1 + Ts/2}, \quad D_{e2} + C_{e2}(sI - A)^{-1}B_d = \rho \frac{s + a}{1 + Ts/2},$$

where

$$\rho \geq \max_{\omega \in \mathbf{R}} \left| \frac{1 + j\omega T}{j\omega + a} \left(e^{-j\omega T} - \frac{1 - j\omega T/2}{1 + j\omega T/2} \right) \right|.$$

The theory of efficient approximation of high (or infinite) order LTI systems by systems of low order (*model reduction*) is an important part of modern control.

Combining equations (1.1) - (1.12) yields a linear model (see Figure 1.3)

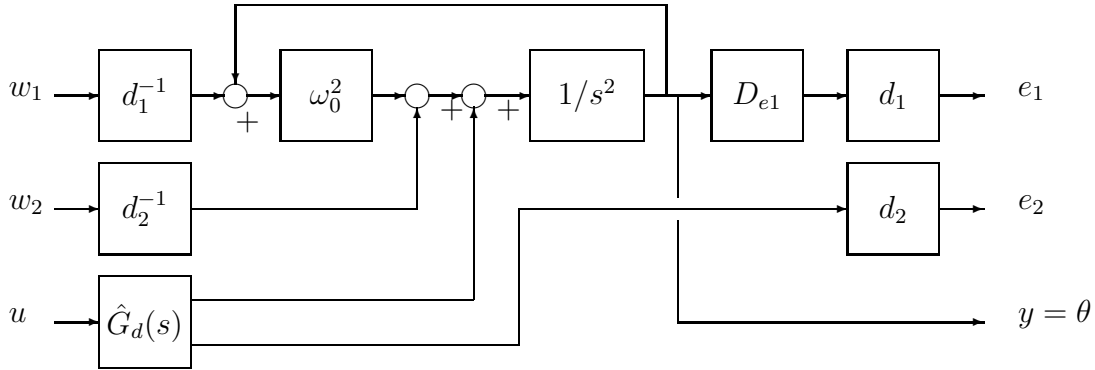


Figure 1.3: Linearized pendulum model

$$\ddot{\theta} - \omega_0^2 \theta = \omega_0^2 d_1^{-1} w_1 + C_d x_d(t) + D_d v(t) + d_2^{-1} w_2, \quad (1.13)$$

$$\dot{x}_d = A_d x_d + B_d v, \quad (1.14)$$

$$e_1 = d_1 D_{e1} \theta, \quad (1.15)$$

$$e_2 = d_2 (C_{e2} x_d + D_{e2} v), \quad (1.16)$$

where L2 gain from $e = [e_1; e_2]$ to $w = [w_1; w_2]$ is known to be not larger than 1. According to the *small gain theorem*, a controller (1.2), (1.3) stabilizes system (1.1) if the L2 gain from w to e in the LTI model (1.2), (1.3), (1.13) - (1.16) is less than 1. The feedback design is now reduced to finding the coefficients of controller (1.2), (1.3) and the positive coefficients d_1, d_2 in (1.13) - (1.16) minimizing the L2 gain.

For $d_1, d_2 > 0$ fixed, one can use the technique of designing LTI feedback control minimizing an L2 gain in an LTI closed loop system, a major component of modern control called *H-Infinity optimization*. Similarly, for a fixed controller (1.2), (1.3), one can use *semidefinite programming* to minimize the L2 gain as a function of scaling parameters d_1, d_2 . Combined together, H-Infinity optimization and semidefinite programming form an ad-hoc procedure of *D-K iteration* for designing robust feedback controllers.

Once a controller (1.2), (1.3) stabilizing (1.1) subject to the assumption $|\theta(t)| \leq \theta_0$ is designed, it remains to calculate the range of initial conditions in (1.1), (1.2), (1.3) which guarantee that the assumption $|\theta(t)| \leq \theta_0$ is satisfied for all $t \geq 0$. This can be achieved by utilizing a variety of L2 gain bounds for the nonlinear closed loop system within the framework of *Integral Quadratic Constraints*.

1.1.2 Efficient MIMO stabilization

Consider a situation where an array of low quality sensors and weak actuators is used to reduce oscillations induced in an undamped flexible system by white noise type disturbances. For simplicity, the flexible system will be represented as a set of m identical point masses moving in a single dimension and connected by identical springs, as shown on Figure 1.4.

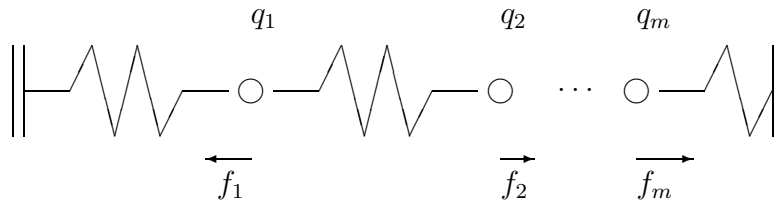


Figure 1.4: MIMO control of a flexible system

System equations are assumed to have the form

$$\begin{aligned}\ddot{q}_1 &= r(q_2 - q_1) + f_1 + bw_1, \\ \ddot{q}_k &= r(q_{k-1} + q_{k+1} - 2q_k) + f_k + bw_k, \quad (k = 2, \dots, m-1) \\ \ddot{q}_m &= r(q_{m-1} - q_m) + f_m + bw_m,\end{aligned}$$

where q_k is the displacement of the k -th mass, f_k is the force produced by the k -th actuator, w_k is the noisy force disturbing the k -th mass, and r, b are constant positive coefficients. In addition, the measurement process is modeled by

$$y_k = g_k + v_k, \quad \dot{g}_k = a(g_k - q_k), \quad (k = 1, \dots, m),$$

where y_k is the k -th mass position measurement, g_k is the signal modeling sensor inertia, and v_k is measurement noise. The objective is to design a feedback law defining $f = [f_1; \dots; f_m]$ as a causal finite order LTI function of measurement $y = [y_1; \dots; y_m]$ which minimizes the mean square value of f in the case when w_i, v_i are modeled as independent white noise stochastic processes.

The practical difficulty of this feedback optimization task is due to the multiplicity of the sensors, as well as the need to minimize the control effort. The problem can be cast as *H2 optimization*, solved efficiently using standard operations of numerical linear algebra.

1.2 Structure of the lectures

The material to be presented can be divided into three categories.

1.2.1 Systems and their approximation

Basic elements of system modeling are introduced here. Signals are considered as functions of continuous time $t \geq 0$. System models of three types are considered.

- (a) *Finite order linear time-invariant* (LTI) systems described by state space equations and, with some loss of information, by transfer matrices. State space LTI models will be the main objects on which optimization and analysis is performed, even when the ultimate objective is feedback design for distributed or nonlinear systems.
- (b) *LTI models defined by transfer matrices* (in general, non-rational), to represent distributed dynamics, and to be approximated by finite order LTI systems.
- (c) *General systems*, defined as arbitrary relations between signals, are used as a generic system class, incorporating LTI models, as well as the error systems produced by approximating general nonlinear and time-varying phenomena by LTI models.

Stability of systems, as well as quality of approximation of one system by another, is defined in terms of L2 gain. The “small L2 gain” criterion is presented for analysis of stability of system interconnections. In particular, sampled data quantized implementations of linear continuous time feedback controllers will also be considered within an approximation and robustness framework.

1.2.2 Explicit optimization of linear systems

H2 and H-Infinity optimization of linear feedback, as well as Hankel optimal model reduction and balanced truncation will be discussed here. The key result is the Kalman-Yakubovich-Popov (KYP) lemma, relating frequency and time domain specifications, establishing conditions of existence of stabilizing solutions of *Riccati equations*, and generalizing Lyapunov theorems to the case of LTI systems with external inputs. Using the KYP lemma and the *generalized Parrot’s lemma* yields explicit formulae for H2 and H-Infinity optimal controllers, computed in terms of Schur decomposition of associated Hamiltonian matrices.

1.2.3 Convex optimization and robustness analysis

Non-classical optimization of linear feedback, as well as advanced analysis of nonlinear, uncertain, and time-varying systems leads to convex optimization (in most situations, semidefinite programming) via *Q-parameterization* and *Integral Quadratic Constraints* (IQC). This can also be combined with H-Infinity optimization, leading to D-K iteration methods in feedback design for robustness.

2 System models

Basic elements of system modeling are introduced here. Signals are considered as functions of continuous time $t \in \mathbf{R}$. System models of three types are considered. *General systems*, defined as arbitrary relations between signals, are used as a generic system class, incorporating *finite order linear time-invariant* (LTI) systems described by state space equations, and *LTI models defined by transfer matrices* (in general, non-rational), to represent distributed dynamics. Stability of systems, as well as quality of approximation of one system by another, is defined in terms of L2 gain. The “small L2 gain” criterion is presented for analysis of stability of system interconnections.

2.1 General systems

General systems are simply sets of signals.

2.1.1 Signals

Signals are defined as locally square integrable functions f of continuous time $t \in \mathbf{R}$.

Definition. An m -dimensional signal is a locally square integrable function $f : \mathbf{R} \mapsto \mathbf{R}^m$. The argument $t \in \mathbf{R}$ is referred to as continuous time. The set of all m -dimensional continuous time signals is denoted by $L_{2e}(\mathbf{R}^m)$.

Recall that two square integrable functions $f_1, f_2 : \mathbf{R} \mapsto \mathbf{R}^m$ are considered equal when their values coincide for *almost all* values of t , i.e. when $f_1(t) = f_2(t)$ on a set with a complement of zero (Lebesgue) measure. (A set $X \subset \mathbf{R}$ is said to have *zero measure* when it can be covered by a countable family of open intervals of arbitrarily small total length.) This corresponds to a *force* interpretation of general continuous time signals: they are expressing themselves through integrals, and an integral of an ordinary function $f = f(t)$ does not change when f is modified on a set of zero measure. Accordingly, for a continuous time signal $f = f(t)$, equality $f = 0$ means that $f(t) = 0$ for almost all t . In this framework, the value $f = f(T)$ of a general continuous time signal f at time T is meaningless, because it can be changed without modifying the signal. To recover the ability to work with individual values of continuous time signals, the notion of an *essential limit*

$$a = \operatorname{ess} \lim_{t \rightarrow T} f(t)$$

can be used, which means that, after modifying the values of f at a set of measure zero,

one can get a signal $\tilde{f}$ such that

$$a = \lim_{t \rightarrow T} \tilde{f}(t).$$

Example. Function $f : \mathbf{R} \mapsto \mathbf{R}$, defined by

$$f(t) = \begin{cases} t^a, & t \neq 0, \\ 0, & t = 0, \end{cases}$$

is a signal for $a > -0.5$. For $a = 0$, it is the same signal as $f(t) \equiv 1$ (in particular, the essential limit of $f(t)$ as $t \rightarrow 0$ equals 1). The Dirac delta $\delta(t)$ is not a signal at all (because it is a *generalized function*, not a locally square integrable function).

Example. As a rule, continuous time signals are represented in MATLAB by column vectors of their uniformly sampled values, as in

$$f_{N,T} = \begin{bmatrix} f(0) \\ f(T) \\ \vdots \\ f(NT) \end{bmatrix},$$

where T is the *sampling time*, and $N + 1$ is the number of samples.

The following code represents and displays a continuous time signal $f(t) = \sin(t)/(1 + t)$ using 200 samples with sampling time 0.1.

```
N=200;           % number of samples
T=0.1;           % sampling time
t=(0:T:(N*T-T/2))'; % first N time samples as a column vector
f=sin(t)./(1+t); % first N samples of f
close(gcf);      % close existing figures, if any
plot(t,f);       % plot using the continuous signal style
grid             % put the grid for convenience
```

Execution of this code produces the plot on Figure 2.5.

A similar result is achieved by using SCILAB's

```
N=200;           // number of samples
T=0.1;           // sampling time
t=(0:T:(N*T-T/2))'; // first N time samples as a column vector
f=sin(t)./(1+t);  // first N samples of f
delete('all');    // close existing graphics
plot(t,f);       // plot using the continuous signal style
xgrid            // put the grid for convenience
```

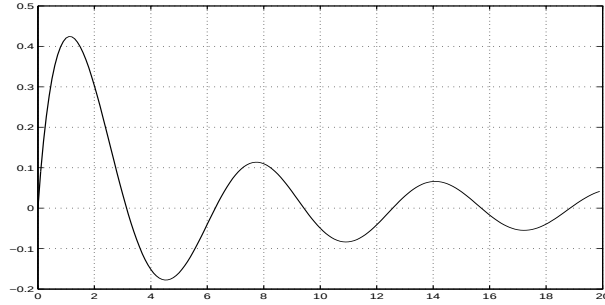


Figure 2.5: MATLAB plot of signal $f(t) = \sin(t)/(1+t)$

We will frequently use the operation of *concatenation on signals*, forming a signal $h = [f; g] \in L_{2e}(\mathbf{R}^{m+q})$ from $f \in L_{2e}(\mathbf{R}^m)$ and $g \in L_{2e}(\mathbf{R}^q)$ according to

$$h(t) = \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}.$$

2.1.2 Behavioral models

An *autonomous system* is simply a set of signals of a fixed dimension, i.e. a subset of $L_{2e}(\mathbf{R}^m)$ for some m . An input/output system with m -dimensional input f and q -dimensional output y is a subset of $L_{2e}(\mathbf{R}^m) \times L_{2e}(\mathbf{R}^q)$, interpreted as the set of all possible input/output pairs (f, y) .

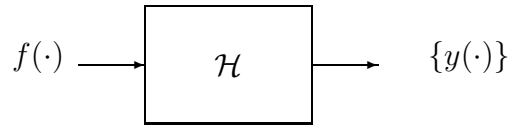


Figure 2.6: System with input f and output y

Example. The set $\mathcal{S} \subset L_{2e}(\mathbf{R}^1)$ of all “unit step” shaped functions

$$f_T(t) = \begin{cases} 1, & t < T, \\ 0, & t \geq T \end{cases}$$

is an autonomous system.

Example. The set of all pairs $(f, y) \in L_{2e}(\mathbf{R}^1) \times L_{2e}(\mathbf{R}^1)$ such that

$$y(t)f(t) = |y(t)|$$

defines the *sign* system $y(t) = \text{sign}(f(t))$: a nonlinear memoryless relation in which some inputs produce multiple outputs.

Example. The *time delay by T* system with m -dimensional input is defined as the the

$$\mathcal{S} = \{(f, y) \in L_{2e}(\mathbf{R}^m) \times L_{2e}(\mathbf{R}^m) : y(t_1) = f(t_2) \text{ for } t_1 - t_2 = T\}.$$

2.1.3 L2 power gain

L2 power gain is a numerical characteristic which is defined for every input/output system. Finiteness of L2 power gain is frequently used as definition of *stability*. As will be supported by the *small gain theorem*, to be stated later, L2 power gain is a key concept in system approximations and robustness analysis.

Definition. L2 power gain of an input/output system $\mathcal{S} \subset L_{2e}(\mathbf{R}^m) \times L_{2e}(\mathbf{R}^q)$ is the maximal lower bound of $\gamma \geq 0$ such that

$$\inf_{T \geq 0} \int_{t=0}^T \{\gamma^2 |f(t)|^2 - |y(t)|^2\} dt > -\infty$$

for every pair $(f, y) \in \mathcal{S}$.

L2 power gain of a system is either a non-negative real number or infinity. For some systems, L2 power gain is the same as the maximal ratio of output and input energies (square integrals) subject to a zero initial condition assumption (frequently called *L2 gain*). For other systems, L2 gain and L2 power gain are different quantities.

The informal rationale behind the definition is as follows: for “zero initial conditions” (whatever this means), we expect the “energy” of the output to be bounded by the energy of the input times the L2 gain squared. Since non-zero initial conditions can produce non-zero output even for zero input, the actual definition says that the difference between the energies must be bounded on one side.

For *memoryless systems*, calculation of L2 power gain is relatively easy.

Theorem 2.1 Let $\phi : \mathbf{R} \times \mathbf{R}^q \mapsto \mathbf{R}^m$ be a piecewise continuous function, such that

$$\sup_{t \in [t_0, t_1], v \neq 0} \frac{|\phi(t, v)|}{|v|} < \infty$$

for all $t_0 < t_1$. The the formula

$$w(t) = \phi(t, v(t))$$

defines a system mapping every signal $v = v(t) \in L_{2e}(\mathbf{R}^q)$ to a signal $w = w(t) = \phi(t, v(t)) \in L_{2e}(\mathbf{R}^m)$. Moreover

(a) L2 power gain of the system is not larger than

$$\lim_{T \rightarrow +\infty} \sup_{t > T, v \neq 0} \frac{|\phi(t, v)|}{|v|};$$

(b) for every $\epsilon > 0$, L2 power gain of the system is not smaller than

$$\lim_{T \rightarrow +\infty} \sup \left\{ k : \text{mes} \left\{ t > T : \sup_{|v| > \epsilon} \frac{|\phi(t, v)|}{|v|} > k \right\} = \infty \right\},$$

where $\text{mes}\{X\}$ denotes the Lebesgue measure of a subset $X \subset \mathbf{R}$.

Example. L2 power gain of system $y(t) = tf(t)$ is infinity. L2 power gain of system $y(t) = e^{-t}f(t)$ is zero. L2 power gain of the *delay by T* transformation introduced earlier is 1 for $T \geq 0$ and ∞ for $T < 0$.

Throughout the class, L2 power gains will be used, sometimes referred to as “L2 gains” for simplicity.

2.1.4 Outer approximations of systems

As a rule, analysis and design for a system which is difficult to handle directly, due to the presence of nonlinearity, time variation, or infinite dimensionality, begins with finding its *outer approximation*.

Definition. Let $\mathcal{S}$ be a system with m -dimensional input and q -dimensional output. A system $\hat{\mathcal{S}}$ with $\hat{m}$ -dimensional input and $\hat{q}$ -dimensional output, where $\hat{m} > m$ and $\hat{q} > q$, is called an *outer approximation* of $\mathcal{S}$ if for every input/output pair $(f, y) \in \mathcal{S}$ there exist $(v, w) \in L_{2e}(\mathbf{R}^{\hat{q}-q}) \times L_{2e}(\mathbf{R}^{\hat{m}-m})$ such that $([f; w], [y; v]) \in \hat{\mathcal{S}}$. Accordingly, the set Δ of all such pairs (v, w) is called the *error system*.

To quantify accuracy of an outer approximation, a simplified description of the error system Δ , such as an L2 power gain bound, can be used.

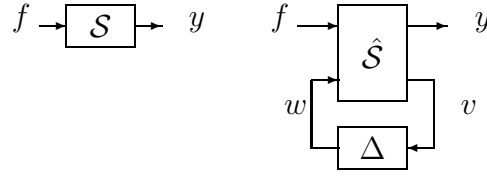


Figure 2.7: Outer approximation

When $\mathcal{S}$ has finite L2 gain, one frequently uses a “direct” outer approximation format defined by

$$\begin{aligned}\hat{\mathcal{S}} &= \{([f; w], [w + y_r; f]) : (f, y_r) \in \mathcal{S}_r, w \in L_{2e}(\mathbf{R}^q)\}, \\ \Delta &= \{(f, y - y_r) : (f, y) \in \mathcal{S}, (f, y_r) \in \mathcal{S}_r\},\end{aligned}$$

where $\mathcal{S}_r \subset L_{2e}(\mathbf{R}^m) \times L_{2e}(\mathbf{R}^q)$ is a system approximating $\mathcal{S}$, as shown on Figure 2.8. However, the direct approach typically does not work with unstable systems.

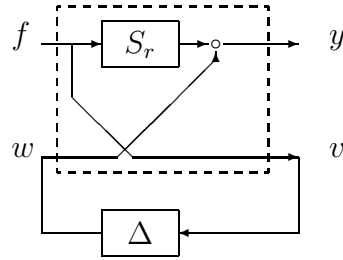


Figure 2.8: “Direct” outer approximation

Example. Nonlinear system

$$\mathcal{S} = \{(f, y) \in L_{2e}(\mathbf{R}) \times L_{2e}(\mathbf{R}) : y(t) = \sin(f(t))\}$$

can be approximated by linear system

$$\hat{\mathcal{S}} = \{([f; w], [y; v]) \in L_{2e}(\mathbf{R}^2) \times L_{2e}(\mathbf{R}^2) : y(t) = af(t) + w(t), v(t) = f(t),$$

with the error system

$$\Delta = \{(v, w) \in L_{2e}(\mathbf{R}) \times L_{2e}(\mathbf{R}) : w(t) = \sin(v(t)) - av(t)\}.$$

The minimal L2 power gain ≈ 0.6086 of Δ is achieved at $a \approx 0.3914$.

Example. Unstable uncertain system

$$\mathcal{S} = \{(f, y) \in L_{2e}(\mathbf{R}) \times L_{2e}(\mathbf{R}) : \dot{y}(t) = ay(t) + f(t), \quad a \in [0, 2]\}$$

can be approximated by linear system

$$\hat{\mathcal{S}} = \{([f; w], [y; v]) \in L_{2e}(\mathbf{R}^2) \times L_{2e}(\mathbf{R}^2) : \dot{y}(t) = y(t) + f(t) + w(t), \quad v(t) = y(t)\},$$

with

$$\Delta = \{(v, w) \in L_{2e}(\mathbf{R}) \times L_{2e}(\mathbf{R}) : w(t) = \delta v(t), \quad \delta = a - 1 \in [-1, 1]\}.$$

The L2 power gain of Δ is not larger than 1. Note that a “direct” approximation of $\mathcal{S}$ by the “central” system

$$\mathcal{S} = \{(f, y) \in L_{2e}(\mathbf{R}) \times L_{2e}(\mathbf{R}) : \dot{y}(t) = y(t) + f(t)\}$$

would lead to an unstable error system (infinite L2 power gain).

2.1.5 Small gain theorem

Let Δ be a system with $\bar{m}$ -dimensional output w and $\bar{q}$ -dimensional input v . Let $\hat{\mathcal{S}}$ be a system with $\hat{m}$ -dimensional input $[f; w]$ and $\hat{q}$ -dimensional output $[y; v]$, where $\hat{m} > \bar{m}$ and $\hat{q} > \bar{q}$. The block diagram on Figure 2.7 defines an interconnection system $\tilde{\mathcal{S}}$ with input f of dimension $m = \hat{m} - \bar{m}$ and output y of dimension $q = \hat{q} - \bar{q}$, defined as the set of all pairs (f, y) for which there exist $(v, w) \in \Delta$ such that $([f; w], [y; v]) \in \hat{\mathcal{S}}$.

When sufficiently good bounds on L2 gains of $\hat{\mathcal{S}}$ and Δ are known, a bound on the L2 gain of $\tilde{\mathcal{S}}$ can be established in the form of the *small L2 gain theorem*.

Theorem 2.2 *If L2 gain of $\hat{\mathcal{S}}$ is not larger than 1 and L2 of Δ is strictly smaller than 1 then L2 gain of $\tilde{\mathcal{S}}$ is not larger than 1.*

The small gain theorem is a tool for proving stability of systems for which outer approximations with small L2 gains of error systems are known.

Proof. Let $\|h\|_T^2$ denote signal energy restricted to a finite interval:

$$\|h\|_T^2 = \int_0^T |h(t)|^2 dt.$$

By assumption, for every $\theta > 1$

$$\inf_{T>0} \{\theta(\|f\|_T^2 + \|w\|_T^2) - \|y\|_T^2 + \|v\|_T^2\} > -\infty,$$

$$\inf_{T>0} \{\|v\|_T^2 - \|w\|_T^2\} > -\infty$$

for all signals f, y, v, w from Figure 2.7. Combining these inequalities yields

$$\inf_{T>0} \{\theta\|f\|_T^2 - \|y\|_T^2\} > 0,$$

which proves the theorem. ■

As a rule, the small gain theorem is used with “scaling”. For example, if it is known that, for the interconnection described by Figure 2.7, L2 gain of $\hat{\mathcal{S}}$ is known to be less than $a > 0$, and L2 gain of Δ is known to be not larger than $b \geq 0$, where $ab < 1$, applying the small gain theorem to the scaled systems

$$\begin{aligned} \hat{\mathcal{S}}_b &= \{([f; w], [y/a; v/a]) : ([f; w], [y; v]) \in \hat{\mathcal{S}}\}, \\ \Delta_b &= \{(v/a, w) : (v, w) \in \Delta\} \end{aligned}$$

yields an L2 power gain bound of a for the closed loop relation between f and y .

2.2 Transfer matrix models of LTI systems

General LTI systems (not necessarily of finite order) are conveniently defined by their transfer matrices, using Laplace transforms.

2.2.1 Laplace transform and continuous time Fourier transform

We will only consider Laplace transforms which regions of convergence containing the region

$$\mathbf{C}_r = \{s \in \mathbf{C} : \operatorname{Re}(s) > r\}$$

for sufficiently large r .

Definition. A function $f : \mathbf{R} \mapsto \mathbf{R}^m$ is said to have Laplace transform with region of convergence $\mathbf{C}_r$ when the integral

$$F(s) = \int_{t=-\infty}^{\infty} f(t)e^{-st}$$

converges absolutely for all $s \in \mathbf{C}$ with $\operatorname{Re}(s) > r$. In this case the function $F(s) : \mathbf{C}_r^{\operatorname{Re}} \mapsto \mathbf{C}^m$ is called Laplace transform of f .

Laplace transform $F : \mathbf{C}_r \mapsto \mathbf{C}^m$ of a function $f : \mathbf{R} \mapsto \mathbf{R}^m$ is *real analytical*, in the sense that $f(s) = w$ implies $f(\bar{s}) = \bar{w}$. However, it is *not true* that all real analytical functions $F : \mathbf{C}_r \mapsto \mathbf{C}^m$ are z -transforms of functions $f : \mathbf{R} \mapsto \mathbf{R}^m$.

Example. $F(s) \equiv 1$ is not Laplace transform of an ordinary function.

The following theorem provides necessary and sufficient conditions for a real analytical function $F : \mathbf{C}_r \mapsto \mathbf{C}^m$ to be Laplace transform of a locally square integrable function $f : \mathbf{R} \mapsto \mathbf{R}^m$.

Theorem 2.3 *Let $F : \mathbf{C}_r^{\operatorname{Re}} \mapsto \mathbf{C}^m$ be a real analytical function. The following conditions are equivalent:*

- (a) *F is Laplace transform of some locally square integrable function $f : \mathbf{R} \mapsto \mathbf{R}^m$;*
- (b) *for every $R > r$ function $\omega \mapsto F(R + j\omega)$ is square integrable on $\mathbf{R}$.*

Moreover, when conditions (a),(b) are satisfied, equality

$$f(t) = \frac{e^{Rt}}{2\pi} \int_{-\infty}^{\infty} F(R + j\omega) e^{j\omega t} d\omega, \quad (2.1)$$

understood as

$$\lim_{\Omega \rightarrow +\infty} \int_{-\infty}^{\infty} \left| f(t) e^{-Rt} - \int_{-\Omega}^{\Omega} F(R + j\omega) e^{j\omega t} d\omega \right|^2 dt = 0,$$

defines a function $f : \mathbf{R} \mapsto \mathbf{R}^m$ for which F is its Laplace transform.

In particular, formula (2.1) implies that a real analytical function $F : \mathbf{C}_r \mapsto \mathbf{C}^m$ is Laplace transforms of a function $f : \mathbf{R} \mapsto \mathbf{R}^m$ satisfying $f(t) = 0$ for $t < T$ if and only if the integrals

$$\int_{-\infty}^{\infty} e^{2TR} |F(R + j\omega)|^2 d\omega$$

are uniformly bounded for $R > R_0$ for some $R_0 \geq r$.

Example. Function $F : \mathbf{C}_0^{\operatorname{Re}} \mapsto \mathbf{C}$, defined by

$$F(s) = \log(s)/(s+1),$$

(using the main branch of the logarithm, i.e. $\log(1) = 0$), is Laplace transform of a signal $f \in L_{2e}(\mathbf{R}^1)$ such that $f(t) = 0$ for $t < 0$.

Laplace transforms are closely related to continuous time Fourier transforms and the Parseval identity.

Theorem 2.4 *Let $f : \mathbf{R} \mapsto \mathbf{R}^m$ be square integrable, i.e. such that*

$$\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty.$$

Then there exists a unique square integrable function $\tilde{f} : \mathbf{R} \mapsto \mathbf{C}^m$ (called the continuous time Fourier transform of f) such that

$$\lim_{\Omega \rightarrow +\infty} \int_{-\infty}^{\infty} \left| f(t) - \int_{-\Omega}^{\Omega} \tilde{f}(\omega) e^{j\omega t} d\omega \right|^2 dt = 0.$$

Moreover, the Parseval identity

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{f}(j\omega)|^2 d\omega \quad (2.2)$$

holds for this function.

The following is an important implication of the Parseval identity.

Theorem 2.5 *Let $F : \mathbf{C}_0^{Re} \mapsto \mathbf{C}^m$ be Laplace transform of a square integrable function $f : \mathbf{R} \mapsto \mathbf{R}^m$. Then*

$$\lim_{r \rightarrow 0} \int_{-\infty}^{\infty} |F(r + j\omega) - \tilde{f}(\omega)|^2 d\omega = 0. \quad (2.3)$$

In other words, continuous time Fourier transform $\tilde{f}$ of a square integrable function $f : \mathbf{R} \mapsto \mathbf{R}^m$ which has Laplace transform in the region $\text{Re}(s) > 0$ is a limit of restrictions of its Laplace transform to the lines $\text{Re}(s) = \text{const}$ approaching the unit circle.

2.2.2 Transfer matrix models of continuous time systems

Recall that a function H is called *meromorphic* on $\mathbf{C}_0$ when $H : \Omega \mapsto \mathbf{C}^{q,m}$ is analytical, where Ω is an open subset of $\mathbf{C}_0$ such that $\mathbf{C}_0 \setminus \Omega$ is countable, and $(s - s_0)^k H(s)$ is bounded in a neighborhood of every point $s_0 \in \mathbf{C}_0 \setminus \Omega$. Note that a meromorphic function $H(s)$ on

$\mathbf{C}_0$ is automatically defined at a point $s_0 \in \mathbf{C}_0$ for which the limit $\lim_{s \rightarrow s_0} H(s)$ exists. For example, functions $H_1(s) = 1/(s-1)$ and $H_2(s) = s/(s-1)$ are meromorphic on $\mathbf{C}_0$ and not defined at $s_0 = 1$, but their difference $H = H_2 - H_1$ is a meromorphic function defined for all $s \in \mathbf{C}_0$.

Definition. A transfer matrix is a meromorphic function defined on $\mathbf{C}_0$ which is bounded in the region $\text{Re}(s) > R$ for some $R > r$.

The following technical result can be used to define general transfer matrix models.

Theorem 2.6 *Let H be a q -by- m transfer matrix. Let $f \in L_{2e}(\mathbf{R}^m)$ be a signal which decays faster than every exponent as $t \rightarrow -\infty$, i.e.*

$$\int_0^\infty e^{rt} |f(-t)|^2 dt < \infty \quad \forall r \in \mathbf{R}.$$

Then there exists a unique signal $y \in L_{2e}(\mathbf{R}^q)$ such that for every $T \in \mathbf{R}$ and $t < T$ the identity $y(t) = y_T(t)$ holds, where y_T is the inverse Laplace transform (with the region of convergence containing $\mathbf{C}_r$ for some $r \in \mathbf{R}$) of $Y_T(s) = H(s)F_T(s)$, and F_T is the Laplace transform of

$$f_T(t) = \begin{cases} f(t), & t < T, \\ 0, & t \geq T. \end{cases}$$

Proof. Note that f_T has Laplace transform F_T in the whole complex plane, and, since H is real analytical in $\mathbf{C}_r$ for some $r > 0$, Y_T is Laplace transform of a function $y_T : \mathbf{R} \mapsto \mathbf{R}^m$. Since $f_T(t) - f_\tau(t) = 0$ for $t < T$ whenever $\tau \geq T$, the integrals

$$\int_{-\infty}^\infty e^{2TR} |F_{g,T}(R + j\omega) - F_{g,\tau}(R + j\omega)|^2 d\omega$$

are bounded as $R \rightarrow \infty$. The boundedness of $H(s)$ implies boundedness of the integrals

$$\int_{-\infty}^\infty e^{2TR} |Y_T(R + j\omega) - Y_\tau(R + j\omega)|^2 d\omega$$

as $R \rightarrow \infty$, which means that $y_T(t) = y_\tau(t)$ whenever $\tau \geq T$ and $t < T$. Hence y is uniquely defined by

$$y(t) = y_{t+1}(t) \quad (t \in \mathbf{R}).$$

■

Definition. Let H be a q -by- m transfer matrix. Transfer matrix model defined by H is the set of all pairs $(f_0, y_0) \in L_{2e}(\mathbf{R}^m) \times L_{2e}(\mathbf{R}^q)$ such that for every $T \in \mathbf{R}$ there exists $f \in L_{2e}(\mathbf{R}^m)$, a signal which decays faster than every exponent as $t \rightarrow -\infty$, and $y \in L_{2e}(\mathbf{R}^q)$, uniquely defined for f by Theorem 2.6, satisfying $f(t) = f_0(t)$ and $y_0(t) = y(t)$ for $t > T$.

Example. When $T \geq 0$, function $H(s) = e^{-Ts}$ is a *transfer matrix*, and hence defines a transfer matrix model. The standard properties of Laplace transforms indicate that the resulting model is identical to the “pure delay by T ” system, defined as the set of all pairs $(f, y) \in L_{2e}(\mathbf{R}^1) \times L_{2e}(\mathbf{R}^1)$ such that $y(t) = f(t - T)$ for all $t \in \mathbf{R}$.

Note that transfer function $H(s) = e^{-s}$ cannot be defined explicitly in MATLAB, because it is not rational. Instead, one can specify an *approximation*, such as $H_e(s) = (1 - 0.5s)/(1 + 0.5s)$, by

```
s=tf('s');
He=(1-0.5*s)/(1+0.5*s)
```

In SCILAB, the corresponding code is

```
s=poly(0,'s');
G=(s+1)/(s^2+3*s+1)
```

2.2.3 H-Infinity norm

For a complex matrix M , let $\|M\|$ denote the *operator norm* of M , i.e. the square root of the largest eigenvalue of $M'M$ (also called the *largest singular number* of M , sometimes denoted by $\sigma_{\max}(M)$).

Theorem 2.7 The H-Infinity norm $\|G\|_\infty$ of transfer matrix G is defined by

$$\|G\|_\infty = \sup_{\operatorname{Re}(s) > 0} \|G(s)\|, \quad (2.4)$$

where the supremum equals infinity when G has a pole in $\mathbf{C}_0$.

When H-Infinity norm is finite, it can be defined in terms of the limit behavior of $\|G(s)\|$ as $\operatorname{Re}(s) \rightarrow +0$.

Theorem 2.8 Assume $G : \mathbf{C}_0 \mapsto \mathbf{C}^{q,m}$ has finite H-Infinity norm. Then $\|G\|_\infty$ equals the maximal possible limit value of $\|G(s)\|$ as $\operatorname{Re}(s) \rightarrow +0$.

Example. H-Infinity norm of $G(s) = e^{-s}$ equals 1. H-Infinity norm of $G(s) = e^s$ equals infinity.

Example. To calculate H-Infinity norms of rational functions in MATLAB, use the `norm.m` function, as in

```
s=tf('s');
norm([1 1/(s+1);1/(s^2+s+1) 0],Inf)
```

In SCILAB, the corresponding code is

```
s=poly(0,'s');
h_norm([1 1/(s+1);1/(s^2+s+1) 0])
```

2.2.4 L2 gain of transfer matrix models

The importance of H-Infinity norm is largely due to the fact that L2 gain of a transfer matrix model equals H-Infinity norm of the transfer matrix.

Theorem 2.9 *L2 gain of a transfer matrix model equals H-Infinity norm of the transfer matrix.*

Proof. To show that L2 gain cannot be larger than H-Infinity norm, use the Parseval theorem. Recall the definition of a transfer matrix model. Consider an input signal $f \in L_{2e}(\mathbf{R}^m)$, assuming, without loss of generality, that

$$f_T(t) = \begin{cases} f(t), & t < T, \\ 0, & t \geq T, \end{cases}$$

has Laplace transform for all $s \in \mathbf{C}$ for all $T \in \mathbf{R}$. Then $y(t) = y_T(t)$ for $t < T$, where the Laplace transforms Y_T and F_T of y_T and f_T respectively are related by $Y_T(s) = H(s)F_T(s)$ for $\text{Re}(s) > 0$. Since f_T is square integrable, the integrals

$$\int_{-\infty}^{\infty} |F_{g,T}(r + j\omega)|^2 d\omega$$

are uniformly bounded for $r > 0$, and, by the boundedness of $H(s)$ for $\text{Re}(s) > 0$, so are the integrals

$$\int_{-\infty}^{\infty} |F_{g,T}(r + j\omega)|^2 d\omega.$$

According to Parseval theorem and the relation between Laplace transforms and continuous time Fourier transforms,

$$\int_{-\infty}^{\infty} \{\gamma^2 |f_T(t)|^2 - |y_T(t)|^2\} dt \geq 0,$$

where γ is the H-Infinity norm of H . Hence

$$\inf_{T>0} \int_0^T \{\gamma^2 |f(t)|^2 - |y(t)|^2\} dt \geq \int_{-\infty}^0 \{\gamma^2 |f(t)|^2 - |y(t)|^2\} dt > -\infty.$$

To show that H-Infinity norm cannot be larger than L2 gain, consider a transfer matrix $H = H(s)$ which defines a transfer matrix model with a finite L2 gain γ .

Let us show first that H has an analytical extension defined on $\mathbf{C}_0$. Indeed, for every vector $f_0 \in \mathbf{R}^m$ consider the inputs

$$f_i(t) = \begin{cases} e^{-t} e_i, & t \geq 0, \\ 0, & t < 0, \end{cases}$$

where e_i for $i = 1, \dots, m$ are the standard basis vectors in $\mathbf{R}^m$. Let y_i be the outputs corresponding to inputs f_i and zero initial conditions, ($g \equiv 0$ in the standard definition). Since f_i are square integrable and L2 gain is assumed to be finite, the outputs y_i are square integrable as well. Hence y_i have Laplace transforms $Y_i = Y_i(s)$, defined for $\text{Re}(s) > 0$, and $Y_i(s) = H(s)F_i(s)$. Therefore $Y(s) = H(s)F(s)$, where Y is the square matrix with columns Y_i and F is the square matrix with columns F_i . Since $F(s) = (s+1)^{-1}I_m$, H has analytical extension defined for $s \in \mathbf{C}_0$ by

$$H(s) = (s+1)Y(s).$$

Since the columns of $Y = Y(s)$ are Laplace transforms of square integrable signals, for every $f_0 \in \mathbf{C}^m$ we have

$$\sup_{r>0} \int_{-\infty}^{\infty} |Y(r+j\omega)f_0|^2 d\omega < \infty.$$

Hence, for an arbitrary $s_0 \in \mathbf{C}_0$,

$$\sup_{r>0} \int_{-\infty}^{\infty} \left| \frac{H(s) - H(s_0)}{s - s_0} f_0 \right|^2 d\omega < \infty$$

as well.

Now for arbitrary $s_0 \in \mathbf{C}_0$ and $f_0 \in \mathbf{C}^m$ let f^c be the complex vector-valued function

$$f(t) = \begin{cases} e^{s_0 t} f_0, & t \geq 0, \\ 0, & t \leq 0. \end{cases}$$

Let f_r, f_i be the signals defined by $f = f_r + jf_i$. Let y_r, y_i be the resulting system responses, defined with zero initial conditions. For $F = F_r + jF_i$ and $Y = Y_r + jY_i$, where F_r, F_i, Y_r, Y_i are the Laplace transforms of f_r, f_i, y_r, y_i , we have

$$Y(s) = Y_0(s) + H(s_0)F(s),$$

where

$$Y_0(s) = \frac{H(s) - H(s_0)}{s - s_0} (1 - e^{(s_0 - s)T}) f_0$$

is Laplace transform of a square integrable function. In other words, the difference $y(t) - H(s_0)f(t)$, where $y = y_r + jy_i$, is a square integrable function. Hence for every

$$\gamma \in (0, |H(s_0)f_0|/|f_0|)$$

the integrals

$$\int_0^T \{\gamma^2 |f_r|^2 + \gamma^2 |f_i|^2 - |y_r|^2 - |y_i|^2\} dt$$

converge to minus infinity as $T \rightarrow \infty$, which implies that at least one of the integrals

$$\int_0^T \{\gamma^2 |f_r(t)|^2 - |y_r(t)|^2\} dt$$

or

$$\int_0^T \{\gamma^2 |f_i(t)|^2 - |y_i(t)|^2\} dt$$

is not bounded from below. ■

Using H-Infinity norm allows one to quantify quality of approximations for LTI systems.

2.2.5 Outer approximations of transfer matrix models

For the purpose of computer-aided analysis and design, non-rational transfer matrix models are approximated by the rational ones. In principle, if H and H_r are two transfer matrices such that H-Infinity norm of $H - H_r$ is small, H_r can be used to define a good approximation of H . However, a bit of caution is need here: the difference of two transfer models $\mathcal{H}_1$ and $\mathcal{H}_2$, defined as the system for which the outputs y , corresponding to a given input f , are differences $y = y_1 - y_2$ of all possible responses y_i of $\mathcal{H}_i$, is *not*, in general, a transfer matrix model.

Example. The difference of two “pure integrator” systems

$$y_i(t) = c_i + \int_0^t f_i(\tau) d\tau$$

responds with an arbitrary constant output to every input $f \in L_{2e}(\mathbf{R}^1)$.

One of many acceptable ways to approximate systems with non-rational transfer matrices is justified by the following statement.

Theorem 2.10 Let H and H_r be two transfer matrices such that H -Infinity norm of $W_1^{-1}(H - H_r)W_2^{-1}$ equals $\gamma < \infty$, where W_1, W_2 are two square transfer matrices such that W_1, W_2, W_1^{-1} and W_2^{-1} have finite H -Infinity norms. Then system with transfer matrix

$$\hat{H} = \begin{bmatrix} H & W_1 \\ W_2 & 0 \end{bmatrix}$$

defines an outer approximation of H for which L_2 gain of the error system equals γ .

The block diagram illustrating Theorem 2.10 is shown on Figure 2.9.

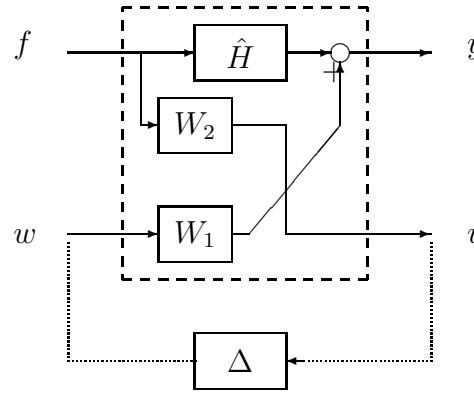


Figure 2.9: Outer approximation of LTI systems

Example. For transfer function

$$H(s) = \frac{e^{-\tau s}}{s - 1},$$

where $\tau > 0$ is a parameter, an approximation in the form

$$H_0(s) = \frac{e^{-\tau}}{s - 1} - \frac{\tau}{1 + 0.5\tau} \frac{1}{1 + 0.5\tau s}$$

can be used. To use the fact that approximation has better quality at low frequencies, use weights $W_1(s) = 1$,

$$W_1(s) = \frac{1 + s}{10 + s}, \quad W_2(s) \equiv 1.$$

L_2 gain of the error system will be given by

$$h = \|(H - H_0)/W\|_\infty,$$

which can be computed using MATLAB code

```

s=tf('s');          % useful "transfer function"
H0=exp(-tau)/(s-1)-(tau/(1+tau/2))/(1+tau*s/2); % approximation of H
W=(1+s)/(1+s/10);    % frequency weight
w=linspace(0,100,10000); % frequency samples
H0w=squeeze(freqresp(H0,w)).'; % frequency response of H0
Ww=squeeze(freqresp(W,w)).'; % frequency response of W0
Hw=exp(-tau*j*w)./(j*w-1); % frequency response of H
h=max(abs((H0w-Hw)./Ww)); % approximate weighted H-Infinity error

```

In particular, for $\tau = 0.1$, the H-Infinity norm of Δ is not larger than 0.0042.

The corresponding SCILAB code is

```

function h=ex2(tau)
s=poly(0,'s');          // useful "transfer function"
j=sqrt(-1);             // square root from -1
H0=exp(-tau)/(s-1)-(tau/(1+tau/2))/(1+tau*s/2); // approximation of H
[A0,B0,C0,D0]=abcd(H0); // state space coefficients of H
W=(1+s)/(1+s/10);       // frequency weight
[Aw,Bw,Cw,Dw]=abcd(W);  // state space coefficients of W
w=linspace(0,100,10000); // frequency samples
H0w=freq(A0,B0,C0,D0,j*w); // frequency response of H0
Ww=freq(Aw,Bw,Cw,Dw,w); // frequency response of W0
Hw=exp(-tau*j*w)./(j*w-1); // frequency response of H
h=max(abs((H0w-Hw)./Ww)); // approximate weighted H-Infinity error

```

2.3 State space models

Finite dimensional state space models are very convenient in system-related calculations.

2.3.1 Definition of state space LTI models

A (continuous time) *state space model* defines a system with m - dimensional input and q - dimensional output, related by

$$y(t) = Cx(t) + Df(t), \quad \dot{x}(t) = Ax(t) + Bf(t) \text{ for } t \geq 0, \quad (2.5)$$

where A, B, C, D are real matrices of dimensions n -by- n , n -by- m , q -by- n , and q -by- m respectively, the differential equation in (2.5) is understood as

$$x(t) = x(0) + \int_0^t \{Ax(\tau) + Bf(\tau)\}d\tau,$$

and $x(0) \in \mathbf{R}^n$ is an arbitrary “initial condition” parameter.

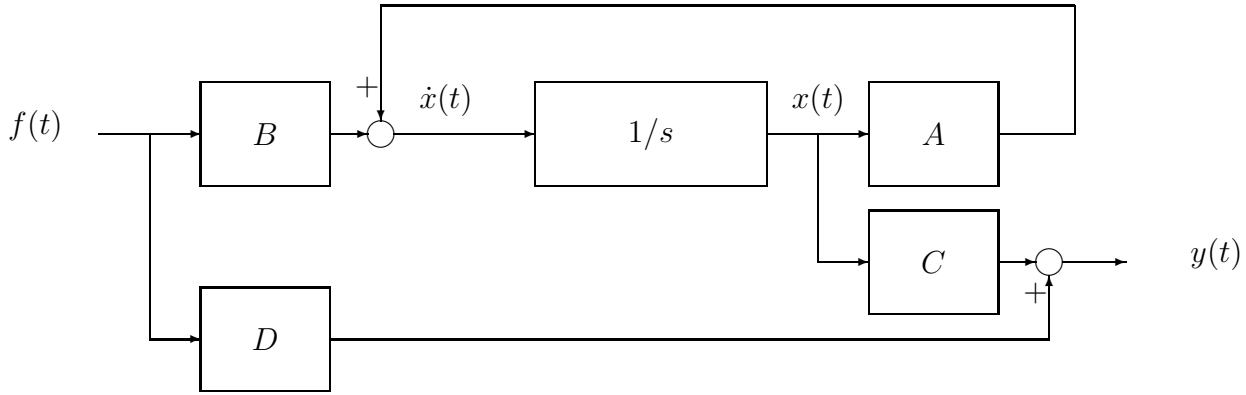


Figure 2.10: A state space model

The special case when $n = 0$, and hence A, B, C are “empty” matrices, generating the *memoryless system* $y(t) = Df(t)$, is also allowed.

A common notation for the system $\mathcal{H}$ defined by (2.5) is

$$\mathcal{H} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right).$$

The *transfer matrix* of state space model (2.5) is the maximal analytical extension of

$$H(s) = D + C(sI - A)^{-1}B, \quad (2.6)$$

defined for all complex s except, possibly, some eigenvalues of A . In addition,

$$H(\infty) \stackrel{\text{def}}{=} \lim_{|s| \rightarrow \infty} H(s) = D.$$

2.3.2 Controllability and observability

The correspondence between state space models, transfer matrices, and transfer matrix models is determined by controllability, observability, and coordinate transformations in the same way as in the discrete time case.

Matrices A, B, C in (2.6) are not uniquely defined by transfer matrix $H = H(s)$.

On one hand, the *change of coordinates* transformation

$$A \mapsto S^{-1}AS, \quad C \mapsto CS, \quad B \mapsto S^{-1}B, \quad (2.7)$$

where S is an arbitrary non-singular n -by- n matrix, has no effect on $H(s)$. This is to be expected, as (2.7) reflects a coordinate change substitution $x := Sx$ in (2.5).

On the other hand, if matrices A, B, C have the block form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}, \quad C = [C_1 \quad C_2], \quad (2.8)$$

then

$$D + C_1(sI - A)^{-1}B_1 = D + C(sI - A)^{-1}B \quad (2.9)$$

for all s but a finite number of them, and hence A_{11}, B_1, C_1, D define same transfer matrix.

The possibility of representing a state space model in the form of (2.8) after a change of coordinates is referred to as lack of *controllability*.

Definition. A pair (A, B) (and also state space model (2.5)) is called *controllable* if there exist no coordinate transformation $A \mapsto S^{-1}AS$, $B \mapsto S^{-1}B$ after which A, B have the form (2.8).

Another opportunity to reduce the dimension of $x(t)$ without changing the transfer matrix appears when

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \quad 0], \quad (2.10)$$

which also implies (2.9).

Definition. The pair (C, A) (and also state space model (2.5)) is called *observable* if there exist no coordinate transformation $A \mapsto S^{-1}AS$, $C \mapsto CS$ after which A, C have the form (2.10).

Clearly, if a state space model is not controllable or not observable, the dimension of x can be reduced without changing the resulting transfer matrix.

Definition. The minimal dimension of $x(t)$ in a state space model with transfer matrix $H = H(s)$ is called *order* of H .

Only a state space model which is *minimal*, i.e. both controllable and observable, can have the number of states which equals the dimension of the corresponding transfer matrix.

Theorem 2.11 All controllable and observable state space models with a given finite order transfer matrix (2.6) have same dimension n .

An LTI system defined by a state space model with transfer matrix H is not necessarily identical to the LTI system defined by the transfer matrix model with the same transfer matrix H .

Example. State space model

$$\dot{x}(t) = x(t), \quad y(t) = x(t) + f(t) \quad (2.11)$$

has transfer matrix $H(z) \equiv 1$. The corresponding transfer matrix model produces a unique output y for every input f , while the state space model produces a continuum of outputs (parameterized by $x(0)$) for each input f .

It can be shown that this disagreement between models is due to a specific lack of controllability.

Theorem 2.12 *A state space model defined by matrices A, B, C, D , where the pair (C, A) is observable, is identical to the transfer matrix model defined by $H(s) = D + C(sI - A)^{-1}B$ if and only if the pair (A, B) is controllable.*

Note that a state space model with transfer matrix $H = H(s)$ is not necessarily the same system as the transfer matrix model defined by H .

Example. State space model

$$\dot{x}(t) = 0, \quad y(t) = x(t) + f(t), \quad (2.12)$$

has transfer matrix $H(s) \equiv 1$. However, its possible response to a zero input is not necessarily zero (as would be in the case of the transfer matrix model defined by $H(s) \equiv 1$): every constant output is also possible.

The following theorem gives a characterization of state space models of minimal dimension defining a given system.

Theorem 2.13 *Among all state space models defining the same system, models with observable pairs (C, A) have the (identical) minimal dimension of the state vector.*

According to the theorem, state space model (2.12) is minimal for the system it defines.

Example. To get a state space model of continuous time system with transfer matrix

$$H(s) = \begin{bmatrix} 1/s & 1/s \\ 1/s & 1/s \end{bmatrix},$$

one can use MATLAB code

```

s=tf('s');
H=[1/s 1/s;1/s 1/s];
H=minreal(ss(H));
[A,B,C,D]=ssdata(H)

```

Note the use of function `minreal.m` which finds a controllable and observable state space model for H .

For this example, SCILAB turns out to be smarter: it produces a minimal realization right away.

```

s=poly(0,'s');
[A,B,C,D]=abcd([1/s 1/s;1/s 1/s])

```

2.3.3 L2 gain of state space models

Numerical calculations of L2 gains of finite order LTI systems are typically performed on state space models. Mathematical results enabling such calculations will be discussed later. This section discusses the relation between L2 gain of a state space model and H-Infinity norm of the corresponding transfer matrix.

Let us call a square matrix a *Hurwitz matrix*, if it does not have eigenvalues s with $\text{Re}(s) \geq 0$.

Definition. A pair (A, B) of real matrices of dimensions n -by- n and n -by- m respectively is called stabilizable if there exists an m -by- n real matrix F such that $A + BF$ is a Hurwitz matrix. A pair (C, A) of real matrices is called detectable when the pair (A', C') is continuous time stabilizable.

Theorem 2.14 *L2 gain of the state space model with coefficients A, B, C, D , where the pair (A, B) is stabilizable and the pair (C, A) is detectable, equals H-Infinity norm of its transfer matrix. In particular, the state space model is stable when A is a Hurwitz matrix.*

Example. State space model (2.12) has infinite L2 gain, despite the fact that its transfer matrix is $H(s) \equiv 1$. The controllability assumption does not hold here.

Example. State space model

$$\dot{x}(t) = 0.5x(t), \quad y(t) = f(t)$$

has L2 gain 1, despite the fact that the pair (A, B) is not stabilizable and the pair (C, A) is not detectable.

3 Classical optimization schemes

Classical formulations of finite order LTI feedback design problems are given in this section.

3.1 Feedback stabilization setup

This section defines the stabilization part of a typical finite order LTI feedback design problem.

3.1.1 Plant, controller, closed loop system

A general objective of feedback design is to provide efficient algorithms for finding causal finite order LTI systems $\mathcal{K}$ to achieve stability and optimize performance in the feedback interconnection shown on Figure 3.11.

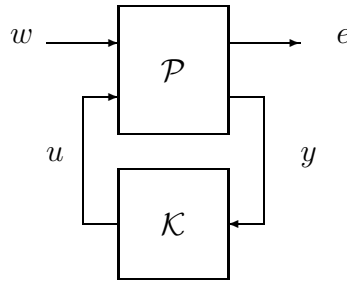


Figure 3.11: Standard feedback design diagram

The elements of the diagram on Figure 3.11 have the following meaning.

$\mathcal{P}$ is the *plant*: an LTI system defined by a state space model, the coefficients of which are known precisely (in some instances, the model can be defined by a proper rational transfer matrix). Input of the plant is partitioned into two components (generally, vectors): *disturbance* w and *control* u . Similarly, output of the plant is partitioned into two components: *cost* e and *measurement* y . Block $\mathcal{K}$ represents a controller: an LTI system defined by a state space or rational transfer matrix model (with input y and output u) to be optimized in design problems, to be analyzed in analysis problems.

The plant equations have the form

$$\dot{x}(t) = Ax(t) + B_1w(t) + B_2u(t), \quad (3.1)$$

$$e(t) = C_1x(t) + D_{11}w(t) + D_{12}u(t), \quad (3.2)$$

$$y(t) = C_2x(t) + D_{21}w(t). \quad (3.3)$$

Note the zero coefficient $D_{22} = 0$ at control in the sensor equation. This assumption about the plant structure is discussed in the next subsection.

The controller equations are chosen among general state space models

$$\dot{x}_f(t) = A_fx_f(t) + B_fy(t), \quad (3.4)$$

$$u(t) = C_fx_f(t) + D_fy(t). \quad (3.5)$$

The resulting closed loop system $\mathcal{G}$ has state space model

$$x_g(t) = A_gx_g(t) + B_gw(t), \quad (3.6)$$

$$e(t) = C_gx_g(t) + C_gw(t), \quad (3.7)$$

where

$$x_g(t) = \begin{bmatrix} x(t) \\ x_f(t) \end{bmatrix}, \quad (3.8)$$

$$\begin{bmatrix} A_g & B_g \\ C_g & D_g \end{bmatrix} = M_0 + M_1 \begin{bmatrix} A_f & B_f \\ C_f & D_f \end{bmatrix} M_2, \quad (3.9)$$

$$M_0 = \begin{bmatrix} A & 0 & B_1 \\ 0 & 0 & 0 \\ C_1 & 0 & D_{11} \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0 & B_2 \\ I & 0 \\ 0 & D_{12} \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & I & 0 \\ C_2 & 0 & D_{21} \end{bmatrix}. \quad (3.10)$$

The feedback interconnection of (3.1)-(3.3) and (3.4),(3.5) is called *stable* if A_g in (3.9) is a Hurwitz matrix. In all problems to be considered, the feedback must be *stabilizing*, i.e. making matrix A_g stable.

3.1.2 On the assumption $D_{22} = 0$

The assumption $D_{22} = 0$ made in (3.1)-(3.3) appears to be justified by “common sense”: since $u(t)$ describes the outcome of some decision-making process within the controller, based on past measurements, causality is expected to prevent $y(t)$ from depending on $u(t)$. At an abstract level, a setup with $D_{22} \neq 0$ is either equivalent to a setup with D_{22} replaced by zero, or is *ill-posed* in a certain sense.

Indeed, when (3.3) is replaced by the more general

$$y = C_2x + D_{21}w + D_{22}u, \quad (3.11)$$

an admissible feedback must satisfy the *well posedness* constraint

$$\det(I - D_f D_{22}) \neq 0. \quad (3.12)$$

However, every controller

$$u = C_f x_f + D_f (C_2 x + D_{21} w + D_{22} u) \quad (3.13)$$

satisfying condition (3.12) can be written as

$$u = \tilde{C}_f x_f + \tilde{D}_f (C_2 x_f + D_{21} w), \quad (3.14)$$

where

$$\tilde{C}_f = (I - D_f D_{22})^{-1} C_f, \quad \tilde{D}_f = (I - D_f D_{22})^{-1} D_f. \quad (3.15)$$

This shows that replacing D_{22} with zero does not reduce the set of achievable closed loop systems. On the other hand, for a general $D_{22} \neq 0$, not every controller (3.14) can be obtained using coefficients defined by (3.15) with some C_f, D_f , because

$$I + \tilde{D}_f D_{22} = (I - \tilde{D}_f D_{22})^{-1}$$

must be an invertible matrix. Therefore, in the case when optimization over controllers (3.14) yields the optimal ones for which $I + \tilde{D}_f D_{22}$ is not invertible, the original feedback setup with (3.13) will be ill-posed, i.e. without an optimal solution.

Example. Consider the task of minimizing H-Infinity norm of the closed loop system for the plant

$$e(t) = y(t) = u(t) + w(t), \quad (3.16)$$

transfer matrix

$$P(s) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Using feedback $u(t) = D_f y(t)$ with $D_f \neq 1$ yields $e(t) = y(t) = (1 - D_f)^{-1} w(t)$. Hence H-Infinity norm of the closed loop system can be made arbitrarily small when $D_f \rightarrow \infty$. However, the zero lower bound is not achieved by any particular admissible controller.

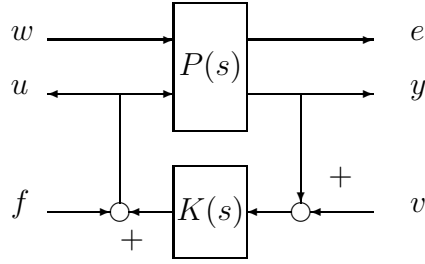


Figure 3.12: Interconnection with interconnection noise

3.1.3 Transfer matrices of feedback interconnection

The complete feedback interconnection system for Figure 3.11 is shown on Figure 3.12, where v and f are interconnection noises.

The resulting interconnection system has inputs w, v, f and outputs e, y, u .

Theorem 3.1 *Let K and P_{22} be the m -by- q transfer matrix of controller $\mathcal{K}$ and the q -by- m lower right corner of the transfer matrix*

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

of plant $\mathcal{P}$. Then transfer matrix H of the interconnection system with input $[w; v; f]$ and output $[e; y; u]$ is given by

$$H = \begin{bmatrix} I & 0 & -P_{12} \\ 0 & I & -P_{22} \\ 0 & -K & I \end{bmatrix}^{-1} \begin{bmatrix} P_{11} & 0 & 0 \\ P_{21} & 0 & 0 \\ 0 & K & I \end{bmatrix}, \quad (3.17)$$

and transfer matrix G of the closed loop system is given by

$$G = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}. \quad (3.18)$$

According to Theorem 3.1, when the pair (A, B) is stabilizable and the pair (C, A) is detectable, stability of interconnection from Figure 3.11 is implied by stability of transfer matrix H from (3.17). On the other hand, under the same assumptions, stability of the closed loop transfer matrix G does not necessarily imply stability of the interconnection.

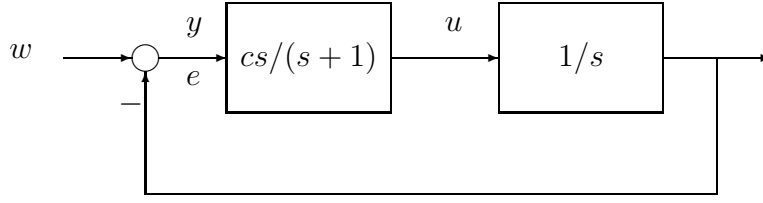


Figure 3.13: Zero/pole cancellation

3.1.4 Zero/pole cancellation: an example

Consider the continuous time SISO (single input, single output) control feedback setup described by the block diagram on Figure 3.13, where $c > 0$ is a constant parameter.

Here $1/s$ describes a servo system to be controlled, and $cs/(s+1)$ is the controller, designed with an intention to minimize lower frequency tracking error (i.e. to make the absolute value of the closed loop transfer function from w to e as small as possible).

This setup corresponds to the general form with

$$P(s) = \begin{bmatrix} 1 & -1/s \\ 1 & -1/s \end{bmatrix}, \quad K(s) = \frac{cs}{s+1}.$$

This results in the closed loop system (input w , output e) with transfer function

$$G(s) = \frac{s+1}{s+1+c}.$$

Since G has no poles with non-negative real part, this gives a (false) impression of stability. Also, as $G(j\omega) \rightarrow 0$ as $\omega \rightarrow \infty$, the tracking objective also appears to be satisfied for large $c > 0$. Nevertheless, the controller does not stabilize the system, and $G(0) = 1$ for every stabilizing controller.

To see that controller $K(s) = cs/(s+1)$ is not stabilizing, introduce the control disturbance input, as shown on Figure 3.14. The transfer function from f to e equals

$$G_{fe}(s) = \frac{s+1}{s(s+1+c)},$$

and hence the interconnection is not stable.

The effect observed in this example can be called *zero/pole cancellation*: cancellation of *unstable* zeros and poles could lead to an instability which does not express itself in the closed loop transfer matrix G .

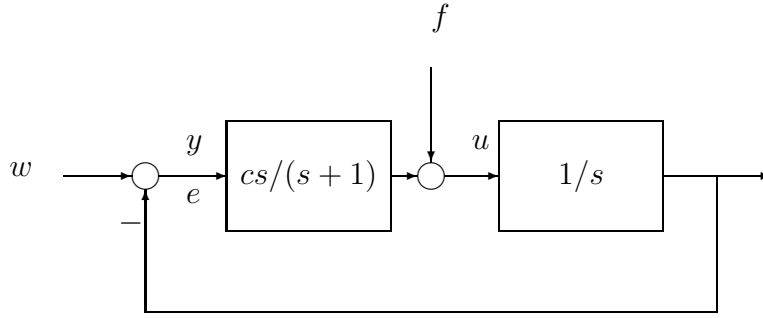


Figure 3.14: Zero/pole cancellation unmasked

3.2 Norm minimization schemes and singularity

In this section we discuss those details of feedback optimization setup which are common between different performance criteria.

A typical feedback optimization setup (see Figure 3.15) calls for minimization of a certain quantity $\Phi = \Phi(\mathcal{G})$, associated with the closed loop system $\mathcal{G}$ (input w , output e), over the set of all controllers $\mathcal{K}$ which make the feedback interconnection with a given plant $\mathcal{P}$ stable.

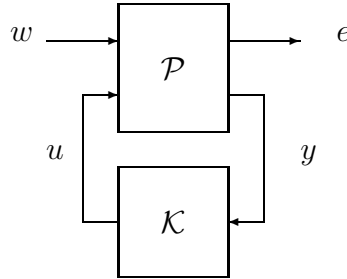


Figure 3.15: Standard feedback design diagram

Among the variety of possible optimization schemes, an important class is formed by *norm minimization*, where the quantity $\Phi = \Phi(G)$ is a *norm* defined on the set of stable transfer matrices, in the sense that

$$\Phi(G) > 0 \text{ for } G \neq 0, \quad \Phi(cG) = |c|\Phi(G) \text{ for } c \in \mathbf{R}, \quad \Phi(G_1 + G_2) \leq \Phi(G_1) + \Phi(G_2).$$

Classical feedback optimization methods provide efficient minimization of H-Infinity norm, H2 norm, and, in a specific setting, Hankel norm, for which norm minimization,

under mild assumptions, is solvable in a finite number of modal decompositions and linear equation solves.

3.2.1 Well posedness of norm minimization problems

Certain conditions of *well posedness* must be satisfied by the plant model for the classical optimization algorithms to work properly. Typically, the following well posedness assumptions are needed to guarantee existence of an optimal controller in a feedback optimization problem.

- (a) The pair (A, B_2) is stabilizable.
- (b) The pair (C_2, A) is detectable.
- (c) The setup is not singular.

Here a state space feedback optimization setup is said to be *control singular* at frequency $\omega \in [0, \infty]$ if matrix

$$E_c(s) = \begin{bmatrix} A - sI & B_2 \\ C_1 & D_{12} \end{bmatrix}$$

is not left invertible at $s = j\omega$ (D_{12} acts as $E_c(j\omega)$ for $\omega = \infty$). The setup is called *sensor singular* at frequency $\omega \in [0, \infty]$ if

$$E_m(s) = \begin{bmatrix} A - sI & B_1 \\ C_2 & D_{21} \end{bmatrix}$$

is not right invertible at $s = j\omega$ (D_{21} acts as $E_m(j\omega)$ for $\omega = \infty$).

As a rule, presence of a singularity reflects certain dangerous omissions/simplifications in the plant model.

3.2.2 Examples involving different types of singularity

Consider the feedback design setup shown on Figure 3.16, where H is a given system to be controlled, K is the feedback system to be designed (accordingly, u is control and y is the measurement), r is either sensor noise or a reference signal modeled by the shaping transfer function W_r , d is the plant disturbance, and $w = [w_1; w_2]$ is the “formal” external input.

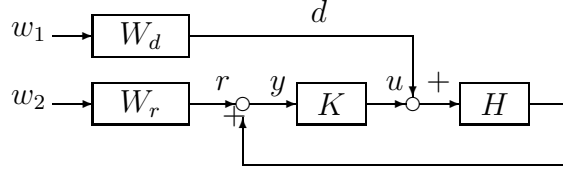


Figure 3.16: Examples of singularity

Control singularity at $\omega = \infty$ The setup with

$$H(s) = \frac{1}{s+1}, \quad W_r(s) = W_d(s) = 1, \quad e = y,$$

apparently aimed at minimizing the reference tracking error, has control singularity at $\omega = \infty$. This is due to the fact that no penalty is put on using high frequency control ($P_{21}(\infty) = 0$). Selecting controller $u = -Ky$, where $K > 0$ is a large positive constant, allows one to make the closed loop gain from w to e arbitrarily small at all frequencies except $\omega = \infty$, but the zero gain cannot be achieved. The “ultimate” controller has an infinite gain.

Control singularity at $\omega = 0$ The setup with

$$H(s) = \frac{1}{s}, \quad W_r(s) = W_d(s) = 1, \quad e = u,$$

apparently aimed at minimizing the control effort, has control singularity at $\omega = 0$. This is due to the fact that no penalty is put on a failure to stabilize the plant. Selecting controller $u = -Ky$, where $K > 0$ is a small positive constant, allows one to make the closed loop gain from w to e arbitrarily small at all frequencies except $\omega = 0$, but the zero gain cannot be achieved. The “ultimate” controller has zero gain, and does not stabilize the system (there will be a zero/pole cancellation at $s = 0$).

Sensor singularity at $\omega = \infty$ The setup with

$$H(s) = W_r(s) = \frac{1}{s+1}, \quad W_d(s) = 1, \quad e = \begin{bmatrix} y \\ u \end{bmatrix},$$

apparently aimed at modeling r as a low frequency reference signal, has sensor singularity at $\omega = \infty$. This is due to the fact that no high frequency noise is assumed to be present in the measurement ($P_{12}(\infty) = 0$). The “ultimate” controller is not proper.

Sensor singularity at $\omega = 0$ The setup with

$$H(s) = \frac{1}{s}, \quad W_r(s) = 1, \quad W_d(s) = 0, \quad e = \begin{bmatrix} y \\ u \end{bmatrix},$$

apparently aimed at ignoring the plant disturbance, has sensor singularity at $\omega = 0$. This is due to the fact that input to a marginally unstable subsystem is assumed to be perfectly known. The “ultimate” controller is not stabilizing (there will be a zero/pole cancellation at $s = 0$), as it attempts to estimate the output of pure integrator by integrating $u(t)$ internally.

3.3 H-Infinity optimization

H-Infinity minimization is a major tool for designing feedback controllers for loop-shaping and robustness in the presence of plant modeling errors.

3.3.1 H-Infinity norm and its calculation

Recall that L-Infinity norm of a rational transfer matrix H is defined by

$$\|H\|_{\infty} = \sup_{\omega \in \mathbf{R}} \|H(j\omega)\|.$$

When H is stable, L-Infinity norm coincides with the H-Infinity norm, i.e. with the L2 gain of the transfer matrix model associated with H . When A in $H(s) = D + C(sI - A)^{-1}B$ is a Hurwitz matrix, L2 gain of the state space model defined by A, B, C, D also equals $\|H\|_{\infty}$.

In MATLAB, function `norm.m` can be used to calculate L-Infinity norm of a rational transfer matrix, as in

```
s=tf('s');
H=[1/(s+1) 1/(s-1);0.3 1];
norm(H,Inf)
```

The corresponding calculation in SCILAB is done using

```
s=poly(0,'s');
H=[1/(s+1) 1/(s-1); 0.3 1];
h_norm(H)
```

A typical algorithm for computing H-Infinity norm is based on the so - called “Kalman - Yakubovich - Popov” (or “Positive real”) lemma, to be discussed later.

3.3.2 Entropy minimization setup

Consider the standard feedback design setup (3.1)-(3.5). Let us say that γ -optimality (in the H-Infinity sense) is *feasible* if there exists a stabilizing controller $\mathcal{K}$ generating a closed loop transfer matrix G with H-Infinity norm less than γ . The *H-Infinity optimization* algorithm, for a given plant $\mathcal{P}$ and real numbers $d, \gamma_{\max}, \gamma_{\min}$, and $\epsilon > 0$ such that $\gamma_{\max} > \gamma_{\min} \geq 0$, begins by finding out whether $\gamma_{\max}$ -optimality is feasible. When it is, the algorithm finds a number $\gamma \in [\gamma_{\min}, \gamma_{\max}]$ such that γ -optimality is feasible, $\gamma - \epsilon < \gamma_{\min}$ or $(\gamma - \epsilon)$ -optimality is not feasible, and computes a controller minimizing the *entropy integral*

$$-\frac{1}{2\pi} \int_{-\infty}^{\infty} \log \det \{I - \gamma^{-2} G(j\omega)' G(j\omega)\} d\omega. \quad (3.19)$$

Note that H-Infinity optimization does not promise a direct calculation of a controller which minimizes the closed loop H-Infinity norm.

Apart from the fact that an “explicit” formula for H-Infinity optimal controller is not known in the general case, there are other good reasons not to seek the absolute optimum in H-Infinity optimization. In particular, when e and w are not scalars, H-Infinity optimal controller is, in general, not unique.

3.3.3 Example: non-unique H-Infinity optimal controller

Consider the H-Infinity feedback optimization setup with two-dimensional signals w, u, e, y , defined by the plant transfer matrix

$$P(s) = \begin{bmatrix} \frac{2}{s+1} & 0 & \frac{s-1}{s+1} & 0 \\ 0 & \frac{1}{s+1} & 0 & \frac{s-1}{s+1} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Since $P_{22} = 0$ in this example, there is no actual feedback loop to close, and hence a feedback controller $K = K(s)$ is stabilizing if and only if it is stable. Accordingly, the closed loop transfer matrix G is given by

$$G(s) = \begin{bmatrix} \frac{2}{s+1} & 0 \\ 0 & \frac{1}{s+1} \end{bmatrix} + \begin{bmatrix} \frac{s-1}{s+1} & 0 \\ 0 & \frac{s-1}{s+1} \end{bmatrix} K(s).$$

Substituting $s = 1$ into the expression for G yields

$$G(1) = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Hence $\|G\|_\infty \geq \|G(1)\| \geq 1$ for all stabilizing controllers K . On the other hand, using

$$K(s) = K_0(s) = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

produces

$$G(s) = G_0(s) = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Since $\|G_0\|_\infty = 1$, K_0 is an H-Infinity optimal controller in this setup. However, K_0 is not the only H-Infinity optimal controller: for example, it is easy to check that

$$K_1(s) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

is optimal as well.

3.3.4 Example: L2 gain optimization via small gain theorem

Consider the task of designing a feedback controller for the standard setup shown on Figure 3.17, where H is an unstable LTI plant with delay,

$$H(s) = \frac{e^{-\tau s}}{s - 1},$$

and K is the controller to be designed to guarantee good tracking of the reference input r at lower frequencies.

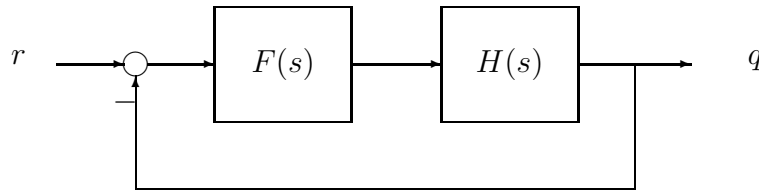


Figure 3.17: Feedback design with delay

We begin by approximating the stable part of H by a lower order transfer function, while bounding the approximation error:

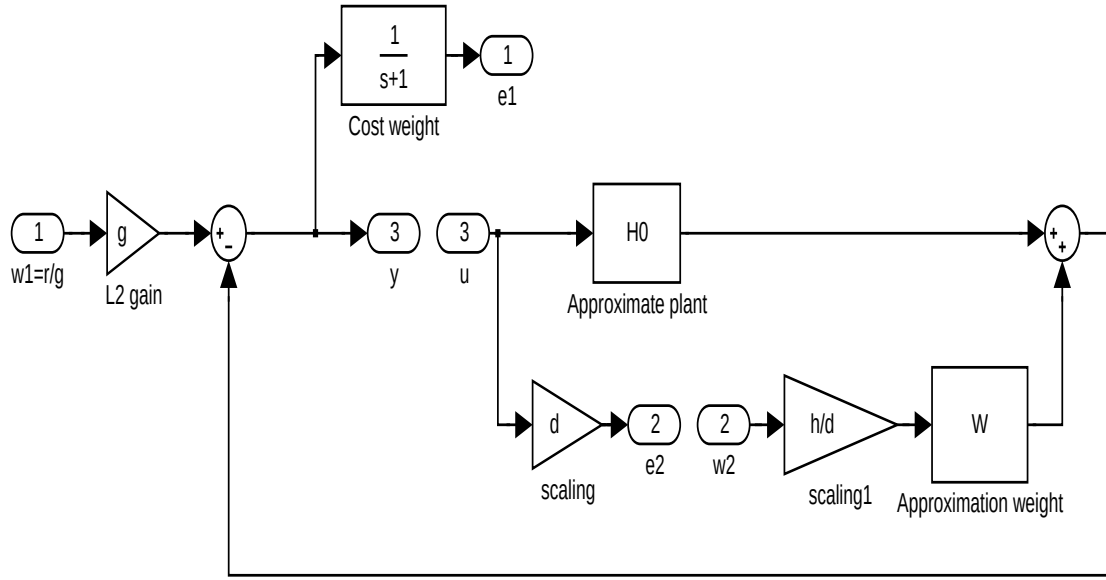
$$H(s) = H_0(s) + hW(s)\Delta(s),$$

where

$$H_0(s) = \frac{e^{-\tau}}{s-1} - \frac{\tau}{1+0.5\tau} \frac{1}{1+0.5\tau s},$$

$$W(s) = \frac{1+s}{10+s}, \quad h = \|(H - H_0)/W\|_{\mathcal{L}\infty},$$

and $\Delta(s)$ is known to have H-Infinity norm not larger than one. The corresponding feedback design diagram and MATLAB code are shown below.



Here γ is the desired guaranteed closed loop L2 gain from the reference input $r = \gamma w_1$ to the frequency weighted tracking error e_1 . Note the way in which the additional scaling parameter d is used. H-Infinity optimization does not provide for an efficient way of selecting the weights, including, in this case, d and γ . This particular implementation relies on manual adjustment of d and γ .

```
function [K,G]=b05_ex7(tau,g,d)
% function [K,G]=b05_ex7(tau,g,d)
%
% L2 gain optimization via small gain theorem
if nargin<1, tau=0.1; end      % default delay value
if nargin<2, g=5; end         % default target L2 gain
if nargin<3, d=1/150; end     % default scaling parameter
```

```

s=tf('s');          % useful "transfer function"
H0=exp(-tau)/(s-1)-(tau/(1+tau/2))/(1+tau*s/2); % approximation of H
W=(1+s)/(1+s/10);    % frequency weight
w=linspace(0,100,10000); % frequency samples
H0w=squeeze(freqresp(H0,w)).'; % frequency response of H0
Ww=squeeze(freqresp(W,w)).'; % frequency response of W0
Hw=exp(-tau*j*w)./(j*w-1); % frequency response of H
h=max(abs((H0w-Hw)./Ww)); % approximate weighted H-Infinity error
disp(['Weighted H-Infinity error: ' num2str(h)])
assignin('base','H0',H0); % workspace assignments
assignin('base','W',W);
assignin('base','g',g);
assignin('base','d',d);
assignin('base','h',h);
P=linmod('b05_ex7mod'); % extract plant model from SIMULINK diagram
p=pck(P.a,P.b,P.c,P.d); % plant model in Mutools format
nmeas=1; % dimension of y
ncon=1; % dimension of u
gmin=0; % lower bound of the binary search interval
gmax=1.1; % upper bound of the binary search interval
tol=0.01; % relative accuracy of binary search
[k,g]=hinfsyn(p,nmeas,ncon,gmin,gmax,tol); % controller optimization
[ak,bk,ck,dk]=unpck(k); % conversion from Mutools format
[ag,bg,cg,dg]=unpck(g);
K=ss(ak,bk,ck,dk);
G=ss(ag,bg,cg,dg);
disp(['Closed loop L2 gain: ' num2str(norm(G,Inf))])

```

3.4 H2 optimization setup

H2 optimization, which includes optimization of steady state Kalman filter gains and LQG (Linear Quadratic Gaussian) feedback optimization, is one of the best developed techniques of LTI system optimization.

3.4.1 H2 norm

The *L2 norm* $\|G\|_2$ of rational transfer matrix G is defined by

$$\|G\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}\{G(j\omega)'G(j\omega)\}d\omega. \quad (3.20)$$

Accordingly $\|G\|_2$ equals infinity when G has poles on the imaginary axis, or $G(\infty) \neq 0$, and is finite otherwise. When G has no poles s with $\text{Re}(s) \geq 0$, the L2 norm will also be called *H2 norm*.

Example. To calculate the L2 norm of $H(z) = 1/(1 + s)$, use the fact that

$$\frac{1}{1 + s} = \int_0^{\infty} e^{-st} e^{-t} dt,$$

and hence

$$\|H\|_2^2 = \int_0^{\infty} e^{-2t} dt = \frac{1}{2}.$$

The same calculation can be done in MATLAB with

```
s=tf('s'); G=1/(1+s); norm(G)^2
```

which returns 0.5000. Remarkably, using

```
s=tf('s'); norm(1/(s-1))
```

produces infinity. The corresponding SCILAB code is

```
s=poly(0,'s');h2norm(1/(1+s))^2
```

(produces an error message when $1/(1 + s)$ is replaced by $1/(1 - s)$).

Calculation of H2 norm is based on the following idea. When

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

where A is a Hurwitz matrix, we have

$$H(s) = \int_0^{\infty} C e^{At} B e^{-st} dt$$

for $\text{Re}(s) \geq 0$, and hence

$$\begin{aligned}\|H\|_2^2 &= \text{trace} \int_0^\infty B'e^{A't}C'Ce^{At}Bdt \\ &= \text{trace}\{B'W_oB\},\end{aligned}$$

where

$$W_o = \int_0^\infty e^{A't}C'Ce^{At}dt$$

is called the *observability Gramian* of the state space model defined by A, B, C . Since

$$e^{A'T}W_o e^{AT} = \int_0^\infty e^{A'(t+T)}C'Ce^{A(t+T)}dt = \int_T^\infty e^{A't}C'Ce^{At}dt$$

for all $T \in \mathbf{R}$, differentiation with respect to T at $T = 0$ yields

$$A'W_o + W_oA = -C'C, \quad (3.21)$$

i.e. the Gramian can be found by solving the linear *continuous time Lyapunov equation*, of which it is a unique solution, provided that A is a Hurwitz matrix.

3.4.2 H2 norms and response to Gaussian white noise

To use stochastic models of signals and systems, we introduce dependence of real variables on an auxiliary parameter $\xi \in (0, 1)$, assuming that every elementary combination of these variables is a *measurable* function of ξ . Accordingly, for a real variable f the quantity

$$\mathbf{E}[|f|] \stackrel{\text{def}}{=} \int_0^1 |f(\xi)|d\xi \in [0, \infty]$$

is well defined, and, assuming $\mathbf{E}[|f|] < \infty$, the expected value

$$\mathbf{E}[f] \stackrel{\text{def}}{=} \int_0^1 f(\xi)d\xi$$

is also well defined.

Let us say that a function $f : (0, 1) \mapsto \mathbf{R}$ has a *Gaussian distribution* with zero mean and variance σ when for every $a \in \mathbf{R}$ the set

$$\{\xi \in (0, 1) : f_c(\xi) < a\}$$

is measurable, and its Lebesgue measure equals

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^a e^{-r^2/2\sigma} dr.$$

An m -dimensional *continuous time Gaussian white noise* $f = f(t)$ is a *linear transformation* which maps every square integrable function $c : [0, \infty) \mapsto \mathbf{R}^m$ to a function $f_c : (0, 1) \mapsto \mathbf{R}$ which has a Gaussian distribution with zero mean and variance

$$\sigma = \int_0^\infty |c(t)|^2 dt.$$

The result of applying f to $c = c(t)$ is expressed symbolically as

$$f_c = \int_0^\infty c(t)' f(t) dt.$$

According to this description, if $y = y(t) = y(t, \xi)$ is response of a stable finite order LTI system $\mathcal{H}$ with a strictly proper transfer matrix H ($H(\infty) = 0$) to a white noise input $f = f(t) = f(t, \xi)$, the expected value of $|y(T)|^2$, is given by

$$\mathbf{E}\{|y(T)|^2\} = \text{trace} \int_0^T h(t)' h(t) dt,$$

where $h = h(t)$ is the unit impulse response of $\mathcal{H}$. When $\mathcal{H}$ is stable, $\mathbf{E}\{|y(T)|^2\}$ converges to the continuous time H2 norm of H .

3.4.3 H2 optimization and its singularity

The *H2 optimization* problem is that of finding, for a given plant $\mathcal{P}$, a controller $\mathcal{K}$ which makes the feedback interconnection well-posed and stable, and minimizes H2 norm of the closed loop transfer matrix G .

One possible interpretation of H2 optimization is that it minimizes sensitivity of output e to a white noise input w .

Subject to the *non-singularity* assumptions about the plant, which are formulated in the same way as in the case of H-Infinity optimization, both discrete and continuous time H2 optimization problems have explicit solutions.

For whatever reason, the routine `h2syn.m`, provided by MATLAB's Mutools for H2 optimization, looks only for strictly proper controllers, and hence does not accept plants

with $D_{11} \neq 0$. Since H2 norm of a state space model with $D \neq 0$ equals infinity, a meaningful controller must satisfy the condition

$$D_{11} + D_{12}D_fD_{21} = 0. \quad (3.22)$$

When $D_{11} \neq 0$, equation (3.22) is not always solvable with respect to D_f . However, in non-singular problems, the solution is unique, when it exists, since D_{12} is left invertible and D_{21} is right invertible. Accordingly, D_f can be defined as

$$D_f = -D_{12}^+D_{11}D_{21}^+,$$

where $D_{12}^+D_{12} = I$ and $D_{21}D_{21}^+ = I$.

Example. Here is a sample continuous time H2 optimization code for plant with state space model

$$\dot{x}(t) = -x(t) + w(t), \quad e(t) = x(t) + u(t), \quad y(t) = w(t),$$

transfer matrix

$$P(s) = \begin{bmatrix} \frac{1}{s+1} & 1 \\ 1 & 0 \end{bmatrix}.$$

```
s=tf('s');           % "s" transfer function
P=[1/(s+1) 1;1 0];   % plant: transfer matrix
% P=[s/(s+1) 1;1 0]; % this would lead to algorithm failure!
[a,b,c,d]=ssdata(P); % plant: state space model
p=pck(a,b,c,d);      % plant: mutools format
[k,g]=h2syn(p,1,1);   % H2 optimization
[ak,bk,ck,dk]=unpck(k); % controller: state space model
tf(ss(ak,bk,ck,dk))   % controller: transfer function
```

Note how changing the plant by uncommenting the third line leads to an algorithm failure (an optimal H2 controller still exists in that case!) due to $D_{11} \neq 0$. The corresponding SCILAB code is

```
s=poly(0,'s');       // "s" transfer function
P=[1/(s+1) 1;1 0];   // plant: transfer matrix
P=[s/(s+1) 1;1 0];   // P will be reset to P=[-1/(s+1) 1;1 0]
[a,b,c,d]=abcd(P);   // plant: state space model
p=syslin('c',a,b,c,d); // plant in lss format
k=lqg(p,[1 1]);      // H2 optimization
k=ss2tf(k)           // controller transfer function
```

3.5 Hankel optimal model reduction setup

Both H-Infinity and H2 optimization setups do not allow the user to impose an arbitrary bound on the order of the optimal controller, though the corresponding solution algorithms are guaranteed to produce a controller of order not larger than order of the plant. Hankel optimal model reduction corresponds to a setup in which $P_{22} = 0$ (no actual feedback loop present), the objective $\Phi = \Phi(G)$ is a weak norm, but, as a compensation, an explicit arbitrary bound can be imposed on the order of the “controller”, which, in this case, represents the reduced model.

3.5.1 The Hankel norm

Let $RH^{q,m}$ denote the set of all q -by- m continuous time stable rational transfer matrices (no poles s with $\text{Re}(s) \geq 0$). Let $RH_-^{q,m}$ denote the set of all q -by- m continuous time anti-stable rational transfer matrices (no poles s with $\text{Re}(s) \leq 0$).

Definition. The Hankel norm $\|G\|_H$ of a q -by- m rational transfer matrix G is defined as

$$\|G\|_H = \inf_{\Delta \in RH_-^{q,m}} \|G - \Delta\|_\infty. \quad (3.23)$$

In other words, Hankel norm of a transfer matrix is its L-Infinity distance to the set of anti-stable transfer matrices.

$\|G\|_H$ equals infinity when G has poles on the imaginary axis, and is finite otherwise. Strictly speaking, the Hankel norm are not a “proper” norm everywhere, as $\|G\|_H = 0$ for a non-zero anti-stable transfer matrix. $\|G\|_H$ is valid as a norm on the subset $RH_0^{q,m}$ of those $G \in RH^{q,m}$ which satisfy $G(\infty) = 0$.

3.5.2 Hankel optimal model reduction

The standard *Hankel optimal model reduction* problem calls for finding, given a stable transfer matrix H and a positive integer r , a *stable* transfer matrix $\hat{H}$ of order less than r for which Hankel norm γ of the difference $H - \hat{H}$ is minimal. In addition, for the optimal $\hat{H}$, an anti-stable transfer matrix Δ can be computed, such that

$$(H - \hat{H} - \Delta)'(H - \hat{H} - \Delta) = \gamma^2 I_m$$

on the imaginary.

As a rule, Hankel optimal reduced models demonstrate high (though not optimal) quality of H-Infinity approximation of the original system. This makes Hankel optimal

model reduction a high quality H-Infinity model reduction algorithm for systems of order not larger than a few thousands.

MATLAB's Mutools offer an implementation of Hankel optimal model reduction for continuous time systems, which uses two operations: "balancing" and actual model reduction, to be explained in the following sections. Here is a sample code that combines the two parts.

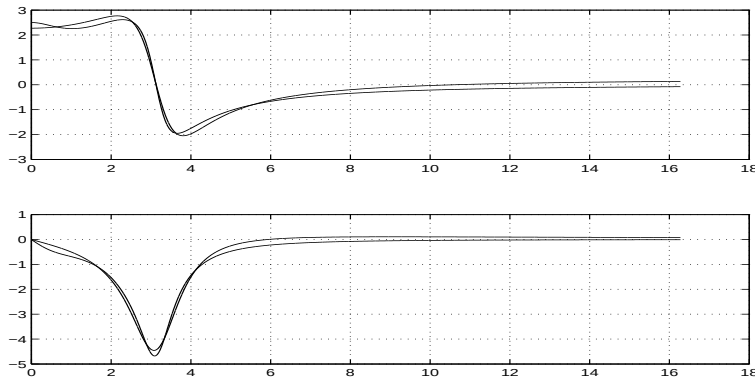


Figure 3.18: Model reduction: frequency responses

```
function Hr=b05_ex10(H,r)
% function Hr=b05_ex10(H,r)
%
% Hankel optimal model reduction

% test inputs
if nargin<1,
    s=tf('s');
    H=1/(s+1)+10/(s^2+s+10)+9/(s^2+2*s+9)-1/(s+2);
end
if nargin<2, r=3; end
% model reduction code
[a,b,c,d]=ssdata(H);           % state space model of H
h=pck(a,b,c,d);               % H in Mutools format
[hbal,sig]=sysbal(h);          % "balancing"
hr=hankmr(hbal,sig,r-1,'d');   % Hankel optimal model reduction
[ar,br,cr,dr]=unpck(hr);       % state space model of reduced system
```

```

Hr=ss(ar,br,cr,dr);
% computing and comparing frequency responses
W=norm(a)/2;
w=linspace(0,W,10000);
Hw=squeeze(freqresp(H,w));
Hrw=squeeze(freqresp(Hr,w));
close(gcf)
subplot(2,1,1); plot(w,real(Hw),w,real(Hrw)); grid
subplot(2,1,2); plot(w,imag(Hw),w,imag(Hrw)); grid

```

For the test example with

$$H(s) = \frac{1}{s+1} + \frac{10}{s^2+s+10} + \frac{9}{s^2+2s+9} - \frac{1}{s+2},$$

frequency responses of H and H_r are compared on Figure 3.18.

4 Q-parameterization and the “waterbed effect”

This lecture discusses some limitations imposed on the closed loop system by the plant equations and by the closed loop stability conditions.

4.1 Q-Parameterization

The so-called “Q-Parameterization” (sometimes referred to as the “Youla Parameterization”): an explicit formula for an arbitrary closed loop transfer matrix generated by a stabilizing controller, expressed as an affine function of an arbitrary stable transfer matrix Q .

4.1.1 The setup

Not every stable transfer matrix $w \rightarrow e$ can be obtained by closing a stabilizing LTI feedback loop $u = K(s)y$ in canonical feedback design setup shown on Figure 4.19. In

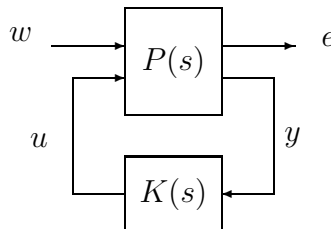


Figure 4.19: LTI feedback

this section, we investigate the limits of LTI feedback’s ability to change the closed loop transfer matrix. More precisely, it will be shown that the set of all closed loop transfer matrices is affine. In addition, the role of unstable poles and zeros of P in limiting LTI feedback designs will be investigated.

Q-parameterization gives a very simple *affine* description of the set of all achievable closed loop transfer matrices as function of the so-called “Q” parameter $Q = Q(s)$, which is an arbitrary stable proper transfer matrix of size $m \times k$, where m is the total number of actuator inputs, and k is the total number of sensors in the system. For design purposes, the result is extremely important: instead of thinking in terms of the controller transfer matrix $K = K(s)$, one will be much better off by designing $Q(s)$.

4.1.2 The Q-parameterization theorem

We consider LTI systems P from Figure 4.19 (where u and y are vectors of size m and k respectively) given in the state space format

$$P(s) := \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{array} \right). \quad (4.1)$$

An LTI feedback controller $u = -Kx$, where

$$K := \left(\begin{array}{c|c} A_f & B_f \\ \hline C_f & D_f \end{array} \right) \quad (4.2)$$

is said to *stabilize* the feedback system if all eigenvalues of the matrix

$$A_{cl} = \begin{bmatrix} A + B_2 D_f C_2 & B_2 C_f \\ B_f C_2 & A_f \end{bmatrix} \quad (4.3)$$

have negative real part. Let $T = T(s)$ be the transfer matrix of the resulting *closed loop system* with input w and output e :

$$T := \left(\begin{array}{cc|c} A + B_2 D_f C_2 & B_2 C_f & B_1 + B_2 D_f D_{21} \\ B_f C_2 & A_f & B_f D_{21} \\ \hline C_1 + D_{12} D_f C_2 & D_{12} C_f & D_{11} + D_{12} D_f D_{21} \end{array} \right). \quad (4.4)$$

Note that the description of T given in (4.4) is very inconvenient for design purposes, because the design parameters A_f, B_f, C_f, D_f enter the transfer matrix defined by (4.4) in a very non-linear way, and are themselves constrained nonlinearly by the condition that A_{cl} in (4.3) must be a Hurwitz matrix. The following result gives a much more useful parameterization of all $T = T(s)$.

Theorem 4.1 *Let F, L be constant gains such that matrices $A + B_2 F$ and $A + L C_2$ are Hurwitz. Then a given transfer matrix $T = T(s)$ can be made equal to the transfer matrix of the closed loop system (4.4) by an appropriate selection of a stabilizing feedback controller (4.2) if and only if*

$$T(s) = T_0(s) + T_1(s)Q(s)T_2(s) \quad (4.5)$$

for some stable proper rational transfer matrix $Q = Q(s)$, where T_0, T_1, T_2 are the transfer matrices of systems

$$T_0 = \left(\begin{array}{cc|c} A & B_2 F & B_1 \\ -L C_2 & A + B_2 F + L C_2 & -L D_{21} \\ \hline C_1 & D_{12} F & D_{11} \end{array} \right), \quad (4.6)$$

$$T_1 = \left(\begin{array}{c|c} \frac{A + B_2 F}{C_1 + D_{12} F} & \frac{B_2}{D_{12}} \end{array} \right), \quad (4.7)$$

$$T_2 = \left(\begin{array}{c|c} \frac{A + LC_2}{C_2} & \frac{B_1 + LD_{21}}{D_{21}} \end{array} \right). \quad (4.8)$$

Moreover, $T(s)$ can be achieved by using a strictly proper stabilizing controller if and only if representation (4.5) with a stable strictly proper rational Q is possible.

Theorem 4.1 is a remarkable result, which, in particular, ensures that any *convex* feedback optimization problem formulated in terms of the closed loop transfer matrices can be solved relatively easily. In (4.5), T_0 represents *some* admissible closed loop design, and the structure of T_1, T_2 imposes limitations on the closed loop transfer matrix.

4.1.3 Derivation of Youla parameterization

Let F, L be defined as in Theorem 4.1, i.e. $A + B_2 F$ and $A + LC_2$ are Hurwitz matrices. For any $u(\cdot)$ let $\hat{x}$ be defined by

$$\dot{\hat{x}} = A\hat{x} + B_2 u + L(C_2 \hat{x} - y).$$

Then for $\delta = x - \hat{x}$ we have

$$\dot{\delta} = (A + LC_2)\delta + (B_1 + LD_{21})w.$$

Let

$$\theta = y - C_2 \hat{x} = C_2 \delta + D_{21} w.$$

Note that θ is the output of system T_2 with input w . Now the equations for $\hat{x}, x_f, u$ can be re-written in the form

$$\begin{aligned} \dot{\hat{x}} &= (A + B_2 D_f C_2) \hat{x} + B_2 C_f x_f + (B_2 D_f - L) \theta, \\ \dot{x}_f &= B_f C_2 \hat{x} + A_f x_f + B_f \theta, \\ u - F \hat{x} &= (D_f C_2 - K) \hat{x} + C_f x_f + D_f \theta. \end{aligned}$$

Hence, for a stabilizing controller (4.2), $u - F \hat{x}$ is the output of the *stable* system

$$Q = \left(\begin{array}{c|c} A_{cl} & \begin{bmatrix} B_2 D_f - L \\ B_f \end{bmatrix} \\ \hline \begin{bmatrix} D_f C_2 - F & C_f \end{bmatrix} & D_f \end{array} \right)$$

with input θ , where A_{cl} is defined by (4.3). Since, in addition, the equations for $x, u - F\hat{x}, w, \delta, e$ can be written as

$$\begin{aligned}\dot{x} &= (A + B_2F)x + B_2(u - F\hat{x}) + (B_1w - B_2F\delta) \\ e &= (C_1 + D_{12}F)x + D_{12}(u - F\hat{x}) + (D_{11}w - D_{12}F\delta),\end{aligned}$$

e equals the output of T_1 with input $u - F\hat{x}$ plus the output of T_0 with input w . This proves that the closed loop transfer matrix from w to e has the form (4.5) with some stable Q . Conversely, if Q is stable, using u which is a sum of $F\hat{x}$ and the output of Q with input θ defines a stabilizing feedback controller $u = Ky$.

Figure 4.20 gives a simple interpretation of Q-parameterization: in order to describe a general form of a stabilizing LTI controller, one has to find a special stabilizing controller first, then use a copy of the stabilized system as an estimator of the sensor output, and feed a stable LTI transformation $Q = Q(s)$ of the difference between the actual and the estimated sensor values back into the control input.

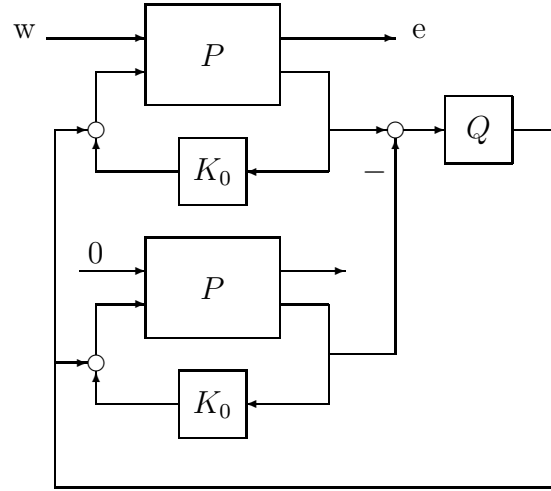


Figure 4.20: Youla parameterization

4.1.4 Open loop zeros

The restrictions on the closed loop transfer function imposed by the Youla parameterization can be understood in terms of the so-called *open loop zeros*.

To define zeros of MIMO systems, consider state space models

$$\dot{x}_1 = ax_1 + bu_1, \quad y_1 = cx_1 + du_1,$$

where the dimensions of x_1, u_1 and y_1 are n, m and k respectively. Consider the complex matrix

$$M(s) = \begin{bmatrix} a - sI & b \\ c & d \end{bmatrix}$$

where $s \in \mathbf{C}$ is a scalar complex variable. The system is said to have a *right zero* at a point s if $\ker M(s) \neq \{0\}$, i.e. if $M(s)$ is not left invertible, or, equivalently, if there exists a non-zero pair (X_1, U_1) of complex vectors such that

$$sX_1 = aX_1 + bU_1, \quad 0 = cX_1 + dU_1.$$

Similarly, the system is said to have a *left zero* at s if the range of $M(s)$ is not the whole vector space $\mathbf{C}^{n+k}$, i.e. if $M(s)$ is not right invertible, or, equivalently, if there exists a non-zero pair (p_1, q_1) of complex vectors such that

$$sp_1 = p_1a + q_1c, \quad 0 = p_1b + q_1d.$$

As it can be seen from the formulae for T_1 and T_2 , the restrictions on the closed loop transfer matrix are caused by the unstable *right zeros* of the system

$$\dot{x} = Ax + B_1w, \quad y = C_2x + D_{21}w, \quad (4.9)$$

and by the unstable *left zeros* of the system

$$\dot{x} = Ax + B_2u, \quad z = C_1x + D_{12}u. \quad (4.10)$$

It can be seen immediately that the right zeros of (4.9) cause problems by obstructing the *observation process*, while the left zeros of (4.10) describe problems which a control action will experience even in the case of a complete knowledge of w and x .

4.1.5 Time-Domain Q-Parameterization

The following statement re-states Theorem 4.1 in a time-domain format.

Theorem 4.2 *A matrix-valued function $Z = Z(t)$ is a closed-loop impulse response from w to z generated in the closed loop system (4.4) by an appropriate selection of a stabilizing feedback controller (4.2) if and only if there exists a function $Q = Q(t)$, elements of which are finite linear combinations of Dirac deltas $\delta = \delta(t)$ and generalized one-sided exponents $t^k e^{st}u(t)$ with $\text{Re}(s) < 0$, such that*

$$Z(t) = (C + D_{12}F)X(t) + D_{12}V(t) + D_{11}\delta(t),$$

where

$$\begin{aligned} V(t) &= \Psi(t)(B_1 + LD_{21}) + Q(t)D_{21}, \\ \dot{\Psi}(t) &= \Psi(t)(A + LC_2) + Q(t)C_2, \\ \dot{X}(t) &= (A + B_2F)X(t) + B_2V(t) + B_1\delta(t), \end{aligned}$$

and $X(\cdot), \Psi(\cdot)$ are zero for $t < 0$.

4.2 The waterbed effect

A common effect, usually associated with unstable zeroes and poles of the open loop plant, makes it theoretically impossible to make certain closed loop transfer functions “small” simultaneously at all frequencies: if amplitude of the frequency response is reduced in one part of the spectrum, it may have to get larger in the other part. This effect, sometimes called the *waterbed effect*, can be explained mathematically in terms of integral inequalities imposed on the closed loop transfer functions. In the basis of such results is the affine characterization of all possible closed loop responses, as well as the Cauchy integral relation for analytical functions.

4.2.1 The Parseval Identity

When working with continuous time systems, the most important integral relation for analytical functions appears to be the *Parseval identity*, which claims that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega)' \tilde{g}(\omega) d\omega = \int_{-\infty}^{\infty} f(t)' g(t) dt, \quad (4.11)$$

where $\tilde{f}, \tilde{g}$ are Fourier transforms of square integrable functions $f, g : \mathbf{R} \mapsto \mathbf{C}^m$.

Let $H_2(\mathbf{R}^m)$ denote the set of all Laplace transforms F of square integrable signals $f : \mathbf{R} \mapsto \mathbf{R}^m$ such that $f(t) = 0$ for $t < 0$. Recall that $F \in H_2(\mathbf{C}^m)$ if and only if F is real analytical, and the integrals

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |F(r + j\omega)|^2 d\omega$$

are uniformly bounded, in which case functions $F_r = F_r(\omega)$, defined by $F_r(\omega) = F(r + j\omega)$, converge in the L2 sense, as $r \rightarrow +0$, to the Fourier transform $\tilde{f}(\omega) \stackrel{\text{def}}{=} F(j\omega)$ of f .

Combining these two statements yields a nice direct integral formula which reconstructs Laplace transform of a square integrable function $f \in L_{2e}(\mathbf{R}^m)$ from the real parts of its values on the imaginary axis.

Theorem 4.3 *If $F \in H_2(\mathbf{R}^m)$ then*

$$F(s) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{Re} F(j\omega) d\omega}{s - j\omega}. \quad (4.12)$$

for all $s \in \mathbf{C}_0$.

Proof. Let F be the Fourier transform of f . Note that f is causal and define $g(t) = f(t) + f(-t)$. Then, by the definition of $F(s)$,

$$F(s) = \int_{-\infty}^{\infty} f(t) e^{-st} dt = \int_{-\infty}^{\infty} g(t) \bar{r}(t) dt,$$

where

$$r(t) = u(t) e^{-\bar{s}t},$$

and $u(\cdot)$ denotes the unit step function. Hence, by the Parseval formula,

$$F(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(j\omega) \bar{R}(j\omega) d\omega = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\operatorname{Re} F(j\omega) d\omega}{s + j\omega},$$

since the Fourier transform $R(j\omega)$ of $r(t)$ is $R(j\omega) = 1/(j\omega + \bar{s})$, the Fourier transform $G(j\omega)$ of g is $G(j\omega) = 2\operatorname{Re} F(j\omega)$, and $2\operatorname{Re} F(j\omega) = 2\operatorname{Re} F(-j\omega)$. ■

Looking separately for the real and imaginary part of $F(s)$, (4.12) yields the *Poisson formula*

$$\operatorname{Re} F(a + jb) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a \operatorname{Re} F(j\omega) d\omega}{a^2 + (b - \omega)^2}. \quad (4.13)$$

Important versions of the Poisson formula are obtained when F is defined as a *logarithm* $F(s) = \log H(s)$ of a minimum-phase stable transfer function.

4.2.2 Integral Identities for Stable Transfer Functions

For the classical SISO feedback setup on Figure 4.21, where $P = P(s)$ is a given strictly proper rational transfer function, and $K(s)$ is an arbitrary proper stabilizing controller, the closed loop sensitivity function

$$S(s) = \frac{1}{1 + P(s)K(s)}$$

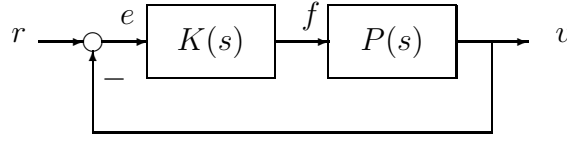


Figure 4.21: A SISO Feedback Setup

must satisfy the interpolation constraints

$$S(p_i) = 0, \quad S(z_i) = 1,$$

where p_i are unstable poles of P , z_i are unstable zeroes of P (including zeros at infinity), and multiplicity counts.

As shown by the Q-parameterization theorem, these conditions are not only necessary but also sufficient for existence of a stabilizing controller producing the closed loop sensitivity function S .

The interpolation conditions can be used to show that the peak values of $|S(j\omega)|$ must be large.

4.2.3 Example: Poisson formula and closed loop bandwidth

Consider the standard SISO feedback design setup shown on Figure 4.21, where $P(s)$ is the given open loop plant model, and $K(s)$ is the stabilizing controller to be designed. We will assume that $P(s)$ has an unstable zero at $s = 2$ and an unstable pole at $s = 3$. According to the classical control, the unstable zero will limit the closed loop bandwidth, in the sense that the closed-loop sensitivity transfer function $S = S(s)$ (from r to e) cannot have small gain on the frequency interval $\omega \in [0, \omega_0]$ when $\omega_0 \gg 2$. Contrary to this, the mathematical theory tells us that, unless P has zeros on the imaginary axis, for every $\epsilon > 0$ and for every $\omega_0 > 0$ there exists a stabilizing controller $C(s)$ such that $|S(j\omega)| < \epsilon$ for all $\omega \in [0, \omega_0]$.

To reconcile these two statements, one can expect that every controller which *theoretically* provides a very large closed loop bandwidth achieves this at the expense of producing very bad behavior at other frequencies. To show that this is indeed the case, let us bound from below the H-Infinity norm of S assuming that $|S(j\omega)|$ does not exceed 0.1 for $|\omega| < 10$ rad/sec.

From the problem formulation we know that S is a stable transfer function, $S(2) = 1$, and $S(3) = 0$. Let

$$B(s) = \frac{s + p_1}{s - p_1} \cdot \frac{s + p_2}{s - p_2} \cdot \dots \cdot \frac{s + p_n}{s - p_n},$$

where p_k are the strictly unstable zeros of $S(s)(s+3)/(s-3)$ (multiplicity counts). Let

$$S_m(s) = S(s)B(s)\frac{s+3}{s-3}.$$

Note that

- (a) $S_m(s)$ does not have strictly unstable zeros or poles;
- (b) $|S_m(j\omega)| = |S(j\omega)|$ for all $\omega \in \mathbf{R}$;
- (c) $|S_m(2)| \geq 5$,

where (a) follows by construction, and (b),(c) are the consequence of the transfer functions $B_p(s) = (s-p)/(s+p)$, with $\text{Re}(p) > 0$, being all-pass, which means $|B_p(j\omega)| = 1$ for $\omega \in \mathbf{R}$ and $|B_p(s)| \leq 1$ for $\text{Re}(s) > 0$.

By (a), the Poisson integral formula can be applied to $\log S_m$, which yields

$$\log |S_m(a+jb)| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{a \log |S_m(j\omega)| d\omega}{(\omega-b)^2 + a^2}$$

for all $a > 0$, $b \in \mathbf{R}$. Setting $a = 2$, $b = 0$, and using (b),(c), we get

$$\log(5) \leq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{2 \log |S(j\omega)| d\omega}{\omega^2 + 4}.$$

Equivalently, the change of the independent variable $\omega := 2\omega$ yields

$$\log(5) \leq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\log |S(2j\omega)| d\omega}{\omega^2 + 1} = \frac{2}{\pi} \int_0^{\infty} \frac{\log |S(2j\omega)| d\omega}{\omega^2 + 1}.$$

Remember that, by the problem formulation,

$$|S(2j\omega)| \leq 0.1 \quad \text{for } \omega < 5.$$

Also let

$$M = \sup_{\omega > 5} |S(2j\omega)|.$$

Taking into account that

$$\int_{\omega_1}^{\omega_2} \frac{d\omega}{1+\omega^2} = \arctan(\omega_2) - \arctan(\omega_1),$$

we have

$$\begin{aligned}
\frac{\pi \log(5)}{2} &\leq \int_0^\infty \frac{\log |S(2j\omega)| d\omega}{\omega^2 + 1} \\
&= \int_0^5 \frac{\log |S(2j\omega)| d\omega}{\omega^2 + 1} + \int_5^\infty \frac{\log |S(2j\omega)| d\omega}{\omega^2 + 1} \\
&\leq \log(0.1) \arctan(5) + \log(M) \left(\frac{\pi}{2} - \arctan(5) \right).
\end{aligned}$$

Hence

$$\log(M) \geq \frac{\pi \log(5)/2 + \arctan(5) \log(10)}{\frac{\pi}{2} - \arctan(5)} \approx 250dB.$$

The calculation was done using MATLAB instruction

$$20 * (\pi * \log_{10}(5) / 2 + \text{atan}(5)) / (\pi / 2 - \text{atan}(5))$$

Indeed, this is a very large *lower bound* of the H-Infinity norm, which makes the design objective $|S(j\omega)| \leq 0.1$ for $\omega \leq 10$ practically infeasible.

5 Kalman-Yakubovich-Popov Lemma

In this section, a variety of results associated with the so-called Kalman - Popov - Yakubovich (KYP, also known as Positive Real or Bounded Real) Lemma are discussed. The “lemma” (which will be presented in the form of several related statements) establishes important relations between state space and frequency domain objects, and plays a key role in deriving state space solutions to H2/H-Infinity feedback optimization and Hankel optimal reduced models. It also enables the semidefinite programming approach to feedback design via convex optimization.

5.1 Motivating examples

This section lists a few major examples which provide motivation for the mathematical constructions related to the KYP Lemma.

5.1.1 Completion of squares in optimal program control

Consider the task of finding a function $f : (0, \infty) \mapsto \mathbf{R}$ for which the integral

$$\Phi(f(\cdot)) = \int_0^\infty (|f(t)|^2 + |x(t)|^2) dt \quad (5.1)$$

is minimal, where

$$\dot{x}(t) = f(t), \quad x(0) = 1. \quad (5.2)$$

This can be viewed as optimization of a pre-calculated control signal in an optimal control problem with perfect information, no disturbances, and asymptotic terminal constraints, though the technique will have far reaching implications.

One can see that minimizing the control effort without regard for $x = x(t)$ leads to $f(t) \equiv 0, x(t) \equiv 1$, i.e. $\Phi = +\infty$. Similarly, minimizing the integral of $|x|^2$ without regard for control, as in

$$f(t) = \begin{bmatrix} -1/e, & 0 < t < \epsilon, \\ 0, & t \geq \epsilon, \end{bmatrix}$$

where $\epsilon \rightarrow +0$, also leads to $\Phi(f(\cdot)) \rightarrow +\infty$.

The trick which allows one to figure out the right balance between control effort and stabilization rate is the so-called *completion of squares*. Note that finiteness of $\Phi(f(\cdot))$ implies that $x(t) \rightarrow 0$ as $t \rightarrow \infty$. Hence, for every constant $p \in \mathbf{R}$,

$$\int_0^\infty 2px(t)f(t)dt = \int_0^\infty \left\{ \frac{d}{dt} p|x(t)|^2 \right\} dt = -p$$

does not depend on $f(\cdot)$. Accordingly,

$$\Phi(f(\cdot)) = p + \int_0^\infty \{|x|^2 + |f|^2 + 2pxu\} dt.$$

Completion of squares means finding p such that the minimum of $|x|^2 + |f|^2 + 2pxf$ with respect to u equals zero, which occurs if and only if

$$|x|^2 + |f|^2 + 2pxu = |f - kx|^2 \quad \forall x, f \in \mathbf{R}$$

for some constant k . Comparing the coefficients at xf and x^2 on both sides of the identity yields $p = -k$ and $1 = k^2$, hence $1 = p^2$. The equation $1 = p^2$ is a special case of (algebraic) *algebraic Riccati equation*, which, in this example, yields a solution of the completion of squares question.

Using the solution $p = 1$ allows one to re-write $\Phi(f(\cdot))$ in the form

$$\Phi(f(\cdot)) = 1 + \int_0^\infty |f + x|^2 dt.$$

This implies $\Phi(f(\cdot)) \geq 1$ for all $f(\cdot)$. Moreover, this suggests that the optimal $f = f(t)$ can be found by solving equation $f(t) = -x(t)$ combined with $dx/dt = f$ and $x(0) = 1$. Indeed, it is straightforward to verify that the corresponding solution $f(t) = -\exp(-t)$, $x(t) = \exp(-t)$ defines the optimal f , for which $\Phi(f) = 1$.

In contrast, the solution $p = -1$ is not as useful for the task at hand. While it does yield a lower bound $\Phi(f) \geq -1$, according to

$$\Phi(f(\cdot)) = -1 + \int_0^\infty |f - x|^2 dt,$$

(subject to the assumption that $x(t) \rightarrow 0$ as $t \rightarrow \infty$), solving $f(t) = x(t)$ combined with $dx/dt = f$ and $x(0) = 1$ defines a function $f = f(t)$ for which $x(t) \rightarrow \infty$ as $t \rightarrow \infty$. The solution $p = 1$ is called *stabilizing*. Apparently, only *stabilizing* solution of the ARE $1 = p^2$ is of some use in solving the optimal program control problem (5.1), (5.2).

Similar arguments apply in the general case of *abstract H_2 optimization*, which is defined as the task of minimizing the functional

$$\Phi(f(\cdot)) = \int_0^\infty \sigma(x(t), f(t)) dt \tag{5.3}$$

subject to the constraints

$$x(0) = x_0, \quad \int_0^\infty (|x(t)|^2 + |f(t)|^2) dt < \infty, \tag{5.4}$$

where x_0 is a fixed, though arbitrary, parameter,

$$\sigma(x, f) = \begin{bmatrix} x \\ f \end{bmatrix}' \Sigma \begin{bmatrix} x \\ f \end{bmatrix}, \quad \Sigma = \begin{bmatrix} G & F' \\ F & R \end{bmatrix} \quad (5.5)$$

is a given Hermitian form with real coefficients, and A, B are given real matrices.

Assuming that $R = R' > 0$, the minimization problem can be solved using completion of squares, which means finding a real symmetric matrix $p = p'$ and a real matrix k such that

$$\sigma(x, f) + 2x'p(Ax + Bf) = |f - kx|_R^2 \quad \forall x, f, \quad (5.6)$$

where $A + Bk$ is a Hurwitz matrix, and $|v|_R^2$ is the shortcut for

$$|v|_R^2 \stackrel{\text{def}}{=} v' R v.$$

Comparing the coefficients on both sides of identity (5.6) yields

$$F + pB = -k'R, \quad G + pA + A'p = k'Rk,$$

which is equivalent to the *algebraic Riccati equation* (ARE)

$$\theta + p\beta + \beta'p = p\theta p \quad (5.7)$$

with

$$\theta = G - FR^{-1}F', \quad \beta = A - BR^{-1}F', \quad \theta = BR^{-1}B'.$$

Moreover, since

$$A + Bk = \beta - \theta p,$$

we are interested in finding a *stabilizing solution* $p = p'$ of ARE (5.7), i.e. such that $\beta - \theta p$ is a Hurwitz matrix. When a stabilizing solution of ARE (5.7) is known, the minimum of $\Phi(u(\cdot))$ equals $|x_0|_p^2$, and is achieved by control satisfying $u = kx$, where $k = -R^{-1}(B'p + F')$.

The completion of squares approach to optimal control poses several fundamental questions.

- (a) Which conditions imposed on A, B, G, F, R guarantee existence of a stabilizing solution of the associated Riccati equation?
- (b) Is it true that a stabilizing solution exists in all well-posed abstract H2 optimization problems?
- (c) Is there an efficient way of computing a stabilizing solution p of ARE (5.7), when such a solution does exist?

These questions, among several other, are answered by a host of results which are generally referred to as the KYP Lemma.

5.1.2 Quadratic storage functions as L2 power gain certificates

Storage functions generalize the notion of a Lyapunov function, providing certificates of input/output behavior of dynamical systems. For example, for the system with input f and output y defined by the ODE model

$$\dot{x}(t) = a(x(t), f(t)), \quad y(t) = c(x(t), f(t)), \quad (5.8)$$

where $a : \mathbf{R}^n \times \mathbf{R}^m \mapsto \mathbf{R}^n$ and $c : \mathbf{R}^n \times \mathbf{R}^m \mapsto \mathbf{R}^k$ are continuous functions, a continuously differentiable function $V : \mathbf{R}^n \mapsto [0, \infty)$ satisfying

$$\gamma^2 |f|^2 - |c(x, f)|^2 - \nabla V(x) a(x, f) \geq 0 \quad \forall x \in \mathbf{R}^n, f \in \mathbf{R}^m, \quad (5.9)$$

where $\nabla V(x)$ denotes the gradient of V at x , proves (“certifies”) that L2 power gain is not larger than $\gamma \geq 0$. Indeed, combining (5.8) with (5.9) yields

$$\gamma^2 |f(t)|^2 - |y(t)|^2 \geq \frac{d}{dt} V(x(t)),$$

which, after integration over $t \in [0, T]$, shows that

$$\int_0^T \{ \gamma^2 |f(t)|^2 - |y(t)|^2 \} dt \geq V(x(T)) - V(x(0)) \geq -V(x(0)).$$

Let us apply storage function analysis to the simple example of LTI system

$$\dot{x}(t) = -x(t) + f(t), \quad y(t) = x(t)$$

($a(x, f) = -x + f$, $c(x, f) = x$, transfer function $1/(s+1)$). For a *quadratic* $V(x) = p|x|^2$, condition (5.9) becomes

$$\gamma^2 |f|^2 - |x|^2 - 2px(-x + f) \geq 0.$$

Since this means positive semidefiniteness of a quadratic form with matrix

$$\begin{bmatrix} 2p-1 & -p \\ -p & \gamma^2 \end{bmatrix},$$

condition (5.9) holds if and only if p satisfies the *Riccati inequality*

$$\gamma^2(2p-1) \geq p^2.$$

The inequality is solvable with respect to p if and only if $\gamma^2 \geq 1$, which is in agreement with the fact that the L2 gain of the system equals 1.

In general, an attempt to find a quadratic storage function $V(x) = x'px$ certifying that L2 power gain of LTI system

$$\dot{x}(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t)$$

is strictly smaller than $\gamma > 0$ leads to the *strict Riccati inequality*

$$\alpha + p\beta + \beta p < p\theta p, \quad (5.10)$$

where

$$\alpha = C'C - C'R^{-1}C, \quad \beta = A - BR^{-1}C, \quad \theta = BR^{-1}B', \quad R = \gamma^2 I - D'D,$$

with the additional constraints $R > 0$ and $p \geq 0$.

The use of quadratic storage functions in LTI system L2 gain verification poses several fundamental questions.

- (a) Which conditions imposed on α, β, θ guarantee existence of a solution $p = p'$ of (5.10)?
- (b) Is it true that (5.10) is solvable whenever the corresponding L2 gain bound is satisfied?
- (c) Is there an efficient way of computing a solution (5.10), when such a solution does exist?

These questions, among several other, are answered by a host of results which are generally referred to as the KYP Lemma.

5.1.3 Differential games

The following example appears to combine the challenges of the previous two. Consider a *differential game*, defined by

$$\dot{x}(t) = u(t) + w(t), \quad e(t) = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}.$$

The objective is to define u as a causal function of x in such a way that the resulting L2 gain from w to e is as small as possible. Informally speaking, the problem is about

finding best strategies for two players, one (“disturbance”) setting $w = w(t)$ and the other (“control”) setting $u = u(t)$, when the objective of control is to keep the integrals

$$\Phi_T = \int_0^T \{\gamma^2 |w(t)|^2 - |x(t)|^2 - |u(t)|^2\} dt$$

bounded from below for $\gamma \geq 0$ as small as possible, while disturbance tries to make Φ_T as small as possible.

The idea of “completing the squares” can be applied in this case, in the form of finding $p \geq 0$ which defines a storage function $V(x) = p|x|^2 \geq 0$ such that

$$\gamma^2 |w(t)|^2 - |x(t)|^2 - |u(t)|^2 - 2px(u + w) = \gamma^2 |w - k_1 x|^2 - |u - k_2 x|^2.$$

If such p, k_1, k_2 do exist, the memoryless feedback controller $u(t) = k_2 x(t)$ guarantees that

$$\Phi_T \geq -p|x(0)|^2 + p|x(t)|^2 + \gamma^2 \int_0^T |w - k_1 x|^2 dt \geq -p|x(0)|^2,$$

no matter what w does. In addition, $w = k_1 x$ appears as the “best strategy” for the disturbance. Solving for p, k_1, k_2 yields a Riccati equation

$$-1 = (\gamma^{-2} - 1)p^2,$$

which has a positive solution if and only if $\gamma > 1$. Moreover,

$$k_2 = -\frac{\gamma}{\sqrt{\gamma^2 - 1}} \rightarrow -\infty \text{ as } \gamma \rightarrow 1 + 0,$$

and hence the closed loop gain can be made arbitrarily close to 1 by using high gain feedback $u = -kx$ where $k \rightarrow +\infty$.

Generalizations of this example lead to the same form of Riccati equations as in (5.7), with the only (though significant) difference that R in (5.5), (5.6) is not a sign definite matrix any longer (though still assumed invertible).

5.2 Systems constructions related to the KYP Lemma

The KYP Lemma reveals a beautiful relation between a number of mathematical objects of importance to control systems.

5.2.1 Riccati equations and inequalities

A Hermitian form

$$h(\eta, \xi) = \begin{bmatrix} \eta \\ \xi \end{bmatrix}' \Sigma_h \begin{bmatrix} \eta \\ \xi \end{bmatrix} = \eta' \alpha \eta + 2\xi' \beta \eta - \xi' \theta \xi, \quad (5.11)$$

where

$$\Sigma_h = \Sigma_h' = \begin{bmatrix} \alpha & \beta' \\ \beta & -\theta \end{bmatrix},$$

and $\alpha = \alpha'$, β , $\theta = \theta'$ are given n -by- n matrices, defines the ARE

$$\alpha + p\beta + \beta'p = p\theta p \quad (p = p') \quad (5.12)$$

with respect to a real symmetric n -by- n matrix $p = p'$. Note that (5.12) is equivalent to

$$h(\eta, p\eta) \equiv 0 \quad \forall \eta \in \mathbf{R}.$$

A solution $p = p'$ of (5.12) is called *stabilizing* when the ODE

$$\dot{\eta} = \left(\frac{dh}{d\xi} \Big|_{\xi=p\eta} \right)$$

is globally asymptotically stable, i.e. when $\beta - \theta p$ is a Hurwitz matrix.

Replacing (5.12) by a matrix inequality yields a *Riccati inequality*, either

$$\alpha + p\beta + \beta'p < p\theta p \quad (5.13)$$

or

$$\alpha + p\beta + \beta'p = p\theta p. \quad (5.14)$$

An important class of Riccati equations and inequalities is generated by a systems-oriented setup, in which

$$h(\eta, \xi) = \tilde{h}(\eta, \xi, \nu) \Big|_{d\tilde{h}/d\nu=0}, \quad (5.15)$$

where

$$\tilde{h}(\eta, \xi, \nu) = \sigma(\eta, \nu) - 2\xi'(a\eta + b\nu),$$

and $\sigma(\eta, \nu)$ is a Hermitian form with real coefficients, such that $R = d^2\sigma/d\nu^2$ is a non-singular matrix. It is easy to verify that a real symmetric matrix $p = p'$ is a stabilizing solution of the corresponding Riccati equation if and only if

$$\sigma(\eta, \nu) - 2\xi'(a\eta + b\nu) = \sigma(0, \nu - k\eta) \quad \forall \nu, \eta,$$

where $a + bk$ is a Hurwitz matrix.

5.2.2 Frequency domain inequalities and spectral factorization

Let $\Pi = \Pi(j\omega)$ be a rational Hermitian matrix-valued function defined on the imaginary axis, satisfying the real symmetry condition

$$\Pi(-j\omega) = \Pi(j\omega)^T \quad \forall \omega \in \mathbf{R}, \quad (5.16)$$

where M^T means the transpose of M (in contrast with M' which means Hermitian conjugation). A question of common interest is verification of *positivity* of Π .

Definition. A rational transfer function $\Pi = \Pi(s)$ with values in $\mathbf{C}^{m,m}$ and no poles on the imaginary axis $j\mathbf{R}$ is called *positive semidefinite on $j\mathbf{R}$* if $\Pi(j\omega) \geq 0$ is Hermitian positive semidefinite for all $\omega \in \mathbf{R}$. Π is called *strictly positive definite* if there exists $\epsilon > 0$ such that

$$\Pi(j\omega) = \Pi(j\omega)' \geq \epsilon I_m \quad \forall \omega \in \mathbf{R}. \quad (5.17)$$

Example. Function $\Pi(s) = \frac{s+2}{s+1}$ is not positive semidefinite on the imaginary axis because its values there are not Hermitian. Function

$$\Pi(j\omega) = \begin{bmatrix} 1 & \frac{1}{1-j\omega} \\ \frac{1}{1+j\omega} & a + \frac{2}{1+\omega^2} \end{bmatrix}, \quad (5.18)$$

where $a \in \mathbf{R}$ is a parameter, is not positive semidefinite for $a < 0$, is positive semidefinite but not strictly positive definite (its smallest eigenvalue converges to zero as $\omega \rightarrow \infty$) for $a = 0$, and is strictly positive definite for $a > 0$.

The following two questions are frequently encountered in control theory and its applications:

- (a) Given a state space models of $\Pi = \Pi(j\omega)$, how to verify that Π is strictly positive definite “analytically”, i.e. without doing an extensive frequency sweep?
- (b) For a strictly positive definite $\Pi = \Pi(j\omega)$, find a rational function Ψ such that both $\Psi(s)$ and $\Psi(s)^{-1}$ are stable transfer matrices, and

$$\Pi(j\omega) = \Psi(j\omega)' \Psi(j\omega) \quad \forall \omega \in \mathbf{R}. \quad (5.19)$$

Representation (5.19) is frequently called *spectral factorization*. It has the following interpretation in the theory of stochastic signals: a transfer matrix $\Pi = \Pi(j\omega)$ satisfying (5.17) is *spectral density* of a stationary random process η , and $\Psi(s)^{-1}$ is transfer matrix of the “whitening” filter, such that passing η through the system with transfer matrix $\Psi(s)^{-1}$ yields a normalized white noise steady state response.

5.2.3 Linear matrix inequalities

Consider the case when $\Pi = \Pi(j\omega)$ in (5.16) is given by

$$\Pi(j\omega) = \begin{bmatrix} (j\omega I - A)^{-1}B \\ I \end{bmatrix}' \begin{bmatrix} G & L \\ L' & R \end{bmatrix} \begin{bmatrix} (j\omega I - A)^{-1}B \\ I \end{bmatrix} \quad (\omega \in \mathbf{R}), \quad (5.20)$$

where $A, B, G = G', L, R = R'$ are known real matrices of dimension n -by- n , n -by- m , and m -by- m respectively, which links Π to ODE state space equations

$$\dot{x}(t) = Ax(t) + Bf(t), \quad (5.21)$$

and Hermitian form

$$\sigma(x, f) = \begin{bmatrix} x \\ f \end{bmatrix}' \begin{bmatrix} G & L \\ L' & R \end{bmatrix} \begin{bmatrix} x \\ f \end{bmatrix}. \quad (5.22)$$

Then for every $f \in \mathbf{C}^m$ and $\omega \in \mathbf{R}$ such that $j\omega$ is not an eigenvalue of A , we have

$$f' \Pi(j\omega) f = \sigma(x, f),$$

where $x \in \mathbf{C}^n$ is defined by

$$j\omega x = Ax + Bf, \quad (5.23)$$

which resembles the “Fourier transform” version of (5.21).

Accordingly, positive definiteness of Π is equivalent to positive definiteness of Hermitian form σ from (5.22) on all subspaces $\mathcal{M}_\omega = \{(x, f)\} \subset \mathbf{C}^n \times \mathbf{C}^m$ defined by (5.23), where $\omega \in \mathbf{R}$ and $j\omega$ is not an eigenvalue of A .

There is an important special class of (generally non-zero) Hermitian forms

$$\sigma_Q(x, f) = 2\text{Re}x'Q(Ax + Bf) = \begin{bmatrix} x \\ f \end{bmatrix}' \Sigma_Q \begin{bmatrix} x \\ f \end{bmatrix}, \quad (5.24)$$

where $Q = Q'$ is a symmetric matrix and

$$\Sigma_Q = \begin{bmatrix} QA + A'Q & QB \\ B'Q & 0 \end{bmatrix}, \quad (5.25)$$

which always have *zero* value subject to (5.23). Note that (5.24) is also an expression for the derivative of the quadratic function

$$V(x(t)) = x(t)'Qx(t) \quad (5.26)$$

subject to (5.21).

Since σ_Q is zero on $\mathcal{M}_\omega$, if $\sigma + \sigma_Q$ is strictly positive definite (or positive semidefinite) for *some* $Q = Q'$ then Π in (5.20) is strictly positive definite (or, respectively, positive semidefinite) as well. Since $\sigma + \sigma_Q$ has a matrix which depends *linearly* on the coefficients G, L, R of σ and on Q , positive definiteness of $\sigma + \sigma_Q$, being equivalent to

$$\tilde{\Sigma}_Q = \begin{bmatrix} G + QA + A'Q & L + QB \\ B'Q + L' & R \end{bmatrix} > 0, \quad (5.27)$$

is referred to as the *Linear Matrix Inequality* (LMI) test of frequency domain positivity of Π .

A solution $Q = Q'$ of LMI (5.27) is an “exact” certificate of (5.17), which eliminates the need for a “frequency sweep” to verify (5.17). In addition, (5.27) allows for some *convex optimization* of G, L, R subject to (5.17). It also defines a quadratic *storage function* (5.26) for the generalized *dissipation inequality*

$$\int_{t_1}^{t_2} \sigma(x(t), f(t)) dt \geq V(x(t_1)) - V(x(t_2)) \quad (5.28)$$

for signals $x \in L_{2e}(\mathbf{R}^n)$, $f \in L_{2e}(\mathbf{R}^m)$ satisfying (5.21).

Example. Transfer matrix (5.18) can be represented in the form (5.20) with

$$A = -1, \quad B = [1 \ 0], \quad G = 1, \quad L = [0 \ 1], \quad R = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix},$$

which corresponds to “state space model” (5.21) and Hermitian form (5.22) given by

$$\dot{x} = -x + f_2, \quad \sigma(x, f) = |x|^2 + 2\operatorname{Re}\bar{x}f_1 + |f_1|^2 + a|f_2|^2.$$

Accordingly,

$$\sigma_Q(x, f) = 2\operatorname{Re}\bar{x}Q(x + f_2),$$

and $\sigma + \sigma_Q$ is the Hermitian form

$$|x|^2 + 2\operatorname{Re}\bar{x}f_1 + |f_1|^2 + a|f_2|^2 + 2\operatorname{Re}\bar{x}Q(-x + f_2),$$

with the coefficient matrix

$$\tilde{\Sigma}_Q = \begin{bmatrix} 1 + 2Q & 1 & Q \\ 1 & 1 & 0 \\ Q & 0 & a \end{bmatrix}.$$

When $a < 0$, $\tilde{S}_Q$ cannot be positive semidefinite. When $a = 0$, using $Q = 0$ makes $\tilde{S}_Q$ positive semidefinite. When $a > 0$, using $Q = a$ makes $\tilde{S}_Q$ strictly positive definite.

5.2.4 Riccati equations and stabilizing solutions

Linear matrix inequality (5.27) implies strict positivity of matrix R . Using a “Schur complement” idea (minimization of Hermitian form $\sigma + \sigma_Q$ with respect to f) on (5.27) yields an equivalent (subject to $R > 0$) inequality

$$\alpha + Q\beta + \beta'Q - Q\theta Q > 0, \quad (5.29)$$

where

$$\alpha = G - LR^{-1}L', \quad \beta = A - BR^{-1}L', \quad \theta = BR^{-1}B'. \quad (5.30)$$

Replacing inequality in (5.29) with equality yields the famous *algebraic Riccati equation* (ARE)

$$\alpha + Q\beta + \beta'Q - Q\theta Q = 0, \quad (5.31)$$

typically to be solved with respect to a real symmetric matrix $Q = Q'$. In general, a n -by- n Riccati equation may have up to 2^n solutions. However, usually only one of them, called *stabilizing*, is of interest in feedback design applications.

Definition. A solution $Q = Q'$ of Riccati equation (5.31) is called stabilizing when $\beta - \theta Q$ is a Hurwitz matrix.

The importance of stabilizing solutions is in part motivated by spectral factorization. It is easy to see that when θ, β, α are defined by (5.30), and $Q = Q'$ is a solution of (5.31), the Hermitian form $\sigma + \sigma_Q$ has the lowest possible (over all $Q = Q'$) rank m , i.e. the identity

$$\sigma(x, f) + 2\operatorname{Re}x'Q(Ax + Bf) = |f - Kx|_R^2 \quad (5.32)$$

holds for all x, f , where $|v|_R^2$ is a shortcut for $v'Rv$, and

$$K = -R^{-1}(B'Q + L') \quad (5.33)$$

and

$$A + BK = \beta - \theta Q,$$

Substituting x, f satisfying (5.23) into (5.32) yields (5.19), where

$$\Psi(s) = R^{1/2}(I - K(sI - A)^{-1}B). \quad (5.34)$$

Transfer matrix Ψ is stable when A is a Hurwitz matrix. Moreover, inverting Ψ yields

$$\Psi(s)^{-1} = (I + K(sI - A - BK)^{-1}B)R^{-1/2}.$$

Therefore Ψ^{-1} is stable when $Q = Q'$ is a stabilizing solution of (5.31).

Example. To compute a spectral factorization of

$$\Pi(j\omega) = \frac{\omega^4 + 4}{(\omega^2 + 1)^2},$$

note first that

$$\Pi(j\omega) = |H(j\omega)|^2, \text{ for } H(s) = \frac{1}{(s+1)^2} \begin{bmatrix} 2 \\ s^2 \end{bmatrix}.$$

Hence, if A, B, C, D are the coefficients of a state space model of $H = H(s)$,

$$\Pi(j\omega) = |D + C(j\omega I - A)^{-1}B|^2,$$

which means that $\Pi(j\omega)$ has a representation (5.20) with

$$\sigma(x, f) = |Cx + Df|^2.$$

For K defined by (5.33), where Q is the stabilizing solution of the corresponding Riccati equation, yields the spectral factor according to (5.34). Here is a MATLAB code doing just that:

```
% b05_ex12.m: spectral factorization of (4+s^4)/(1-s^2)^2
s=tf('s'); H=[2;s^2]/(1+s)^2; % auxiliary transfer matrix
[A,B,C,D]=ssdata(H); % A,B
G=C'*C; L=C'*D; R=D'*D; % coefficients of sigma
t=G-L*inv(R)*L'; b=A-B*inv(R)*L'; a=B*inv(R)*B'; % ARE coefficients
Q=are(b,a,t); % solving ARE
K=-inv(R)*(B'*Q+L'); % optimal feedback gain
tf(sqrtm(R)*ss(A,B,-K,1)) % spectral factor
```

5.2.5 Stable invariant subspaces of Hamiltonian matrices

A direct calculation shows that every solution $Q = Q'$ of ARE (5.31) satisfies

$$H \begin{bmatrix} I \\ -Q \end{bmatrix} = \begin{bmatrix} I \\ -Q \end{bmatrix} (\beta - \theta Q), \quad H' \begin{bmatrix} Q \\ I \end{bmatrix} = - \begin{bmatrix} Q \\ I \end{bmatrix} (\beta - \theta Q), \quad (5.35)$$

where

$$H = \begin{bmatrix} \beta & \theta \\ \alpha & -\beta' \end{bmatrix} \quad (5.36)$$

is the *Hamiltonian matrix* associated with ARE (5.31). Hence the columns of $[I; -Q]$ span a subspace in $\mathbf{R}^{2n}$ which is H -invariant, and the restriction of H on the subspace

is similar to $\beta - \theta Q$ (in particular, has same set of eigenvalues). Similarly, the columns of $[Q; I]$ span a subspace in $\mathbf{R}^{2n}$ which is H' -invariant, and the restriction of H' on the subspace is similar to $-(\beta - \theta Q)$.

Now assume that $Q = Q'$ is a stabilizing solution of ARE (5.31), i.e. $\beta - \theta Q$ is a Hurwitz matrix. Then the columns of $[I; -Q]$ span the *stable invariant subspace* $\mathcal{H}_+$ of H , and the columns of $[Q; I]$ span an anti-stable invariant subspace $\mathcal{H}_-$ of H . Therefore, for a stabilizing solution Q of (5.31) to exist, the Hamiltonian matrix H in (5.36) must have no eigenvalues on the imaginary axis. Moreover, Q can be found by forming a real $2n$ -by- n matrix $U = [U_x; U_\psi]$, where U_x and U_ψ are n -by- n , such that the columns of U form a basis in $\mathcal{H}_+$, and defining $Q = -U_\psi U_x^{-1}$.

In general, a Hamiltonian matrix (5.36) with $\alpha = \alpha'$ and $\theta = \theta'$ has spectrum “symmetric” with respect to the imaginary axis, because, due to

$$J^{-1}HJ = -H', \quad \text{where } J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix},$$

H is similar to minus its own Hermitian conjugation. Therefore, if H has no eigenvalues on the imaginary axis, it will have an n -dimensional stable invariant subspace and n -dimensional anti-stable invariant subspace. This, however, does not necessarily guarantee existence of a stabilizing solution of the associated Riccati equation, as the corresponding matrix U_x could be singular.

A MATLAB implementation of ARE solver can be constructed using functions `schur` and `schord`:

```
function Q=b05_ex13(a,b,c)
% function Q=b05_ex13(a,b,c)
%
% a rough implementation of ARE
n=size(a,1); % basic dimension
H=[a b;c -a']; % Hamiltonian matrix
[U,T]=schur(H,'complex'); % Schur decomposition
[U,T]=schord(U,T,real(diag(T))); % ordering eigenvalues
Ux=[real(U(1:n,1:n)) imag(U(1:n,1:n))]; % Ux
Upsi=[real(U(n+1:2*n,1:n)) imag(U(n+1:2*n,1:n))]; % Upsi
Q=-Upsi/Ux; % form Q
mr=max(real(eig(a-b*Q))); % is answer reasonable?
ep=max(abs(c+Q*a+a'*Q-Q*b*Q));
if ((mr>=0)|(ep>1e-6)), Q=[]; end % discard unreasonable
```

5.2.6 Abstract H2 optimization and completion of squares

The *Abstract H2 optimization* is the task of finding square integrable signals $f \in L_{2e}(\mathbf{R}^m)$, $x \in L_{2e}(\mathbf{R}^n)$ satisfying the differential equation (5.21) with additional constraints

$$x(0) = x_0, \quad \int_0^\infty (|x(t)|^2 + |f(t)|^2) dt < \infty, \quad (5.37)$$

where x_0 is a fixed, though arbitrary, parameter, and minimizing, subject to these restrictions, the integral

$$J(f(\cdot), x(\cdot)) = \int_0^\infty \sigma(x(t), f(t)) dt \rightarrow \min. \quad (5.38)$$

The minimization problem (5.38) is easy to solve when a stabilizing solution $Q = Q'$ of ARE (5.31) with θ, β, α defined by (5.30), where $R > 0$, is available. Indeed, for *every* solution $Q = Q'$ of the ARE, integrating the implied identity (5.32) yields

$$\int_0^\infty \sigma(x(t), f(t)) dt = x_0' Q x_0 + \int_0^\infty |f(t) - Kx(t)|_R^2 dt. \quad (5.39)$$

Representation (5.39) means that the value of J can never be less than $x_0' Q x_0$. This lower bound could be achieved only when

$$f(t) = Kx(t). \quad (5.40)$$

Combining (5.40) with (5.21) yields an explicit expression for the optimal pair $f(\cdot), x(\cdot)$:

$$x(t) = e^{(A+BK)t} x_0, \quad f(t) = K e^{(A+BK)t} x_0,$$

which is square integrable when $A + BK$ is a Hurwitz matrix, i.e. when $Q = Q'$ is a *stabilizing* solution. It is also easy to see that, when θ, β, α are defined by (5.30) with $R = R' > 0$, the stabilizing solution $Q = Q_{st}$ is greater or equal than any other possible solution $Q = Q_{un}$ of the (5.31), because *every* solution provides a lower bound for $J(f(\cdot), x(\cdot))$ via (5.39).

There exists an important connection between the Hamiltonian matrix H from (5.36) and the abstract H2 optimization, via the standard first order *optimality conditions* from the calculus of variations. The corresponding result says that the pair of signals $f \in L_{2e}(\mathbf{R}^m)$, $x \in L_{2e}(\mathbf{R}^n)$ is optimal only if there exists a square integrable signal $\psi \in L_{2e}(\mathbf{R}^n)$ (the “dual variable”) such that

$$\dot{\psi}(t) = Gx(t) - A'\psi(t) + Lf(t), \quad B'\psi(t) = L'x(t) + Rf(t). \quad (5.41)$$

Using the second equation in (5.41) to eliminate $f(t)$ from (5.21) and from the first equation in (5.41) yields the *Hamiltonian system*

$$\frac{d}{dt} \begin{bmatrix} x(t) \\ \psi(t) \end{bmatrix} = H \begin{bmatrix} x(t) \\ \psi(t) \end{bmatrix}, \quad (5.42)$$

where H is the Hamiltonian matrix from (5.36).

Example. A simple example of abstract H2 optimization is given by

$$A = 0, \quad B = 1, \quad x_0 = 1, \quad G = 1, \quad L = 0, \quad R = 1,$$

i.e.

$$\dot{x} = u, \quad x(0) = 1, \quad J(u) = \int_0^\infty (|x|^2 + |f|^2) dt \rightarrow \min.$$

The corresponding ARE has coefficients

$$\theta = 1, \quad \beta = 0, \quad \alpha = 1,$$

i.e. is the simple quadratic equation $Q^2 = 1$. It has two solutions $Q = \pm 1$, of which the largest ($Q = 1$) is the stabilizing one. The corresponding “optimal gain” K equals -1 , which means that

$$x(t) = e^{-t}, \quad f(t) = -e^{-t}$$

is the optimal pair. The Hamiltonian matrix

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

does not have eigenvalues on the imaginary axis, and its stable invariant subspace is spanned by vector

$$\begin{bmatrix} U_x \\ U_\psi \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -Q \end{bmatrix}.$$

5.3 Main statements of the KYP lemma

This section contains the formulations and proofs of the main statements associated with the KYP lemma.

5.3.1 Checking generalized positivity

The following is an “abstract” version of the positivity verification statement of KYP lemma.

Theorem 5.1 *Let M, N be two real q -by- p matrices. Let $P = P'$ be a symmetric p -by- p matrix. The following conditions are equivalent:*

- (a) $u'Pu > 0$ for all $u \in \mathbf{C}^p$ such that $u \neq 0$ and $Mu = zNu$ for some $z \in \mathbf{C}$, $|z| = 1$;
- (b) there exists q -by- q real symmetric matrix $Q = Q'$ such that $P > M'QM - N'QN$.

Proof: (b) $\Rightarrow$ (a). Since $P > M'QM - N'QN$, we have $u'Pu > u'M'PMu - u'N'PNu$ for all $u \neq 0$. If, in addition, $Mu = zNu$ and $|z| = 1$ then $u'M'PMu = u'N'PNu$. Hence $u'Pu > 0$.

Proof: (a) $\Rightarrow$ (b). Assume that (b) is not true. Then the convex set of p -by- p symmetric matrices

$$\Omega = \{P - M'QM + N'QN - \Delta : \Delta = \Delta' > 0, Q = Q'\}$$

does not contain zero. Hence there exists a linear functional τ (not identically equal to zero) on the vector space of symmetric matrices, $S \mapsto \tau(S) = \text{trace}(ZS)$, where $Z = Z' \neq 0$ is a symmetric real p -by- p matrix, such that τ is non-positive on Ω , i.e.

$$\text{trace } Z(P - M'QM + N'QN - \Delta) \leq 0 \quad \forall \Delta = \Delta' > 0, Q = Q'. \quad (5.43)$$

Since $\Delta = \Delta' > 0$ is arbitrary, Z must be positive semidefinite, i.e. $Z = XX'$ for some real p -by- p matrix $X \neq 0$. Using (5.43) with $\Delta \rightarrow 0$ yields

$$\text{trace}(ZP) + \text{trace } Q(NZN' - MZM') \leq 0 \quad \forall Q = Q',$$

which implies $NZN' = MZM'$, or $NXX'N' = MXX'M'$. If two real matrices L_1, L_2 satisfy $L_1L_1' = L_2L_2'$ then there exists a real matrix U which is orthogonal (i.e. $UU' = U'U = I$) and such that $UL_1 = L_2$. Applied to $L_1 = NX$, $L_2 = MX$, this yields $NXU = MX$. Let $\{f_k\}_{k=1}^p$ be the orthonormal family of (possibly complex) eigenvectors of U , i.e.

$$Uf_k = z_k f_k, \quad \sum_{k=1}^p f_k f_k' = I_p.$$

Then, for $u_k = Xf_k$,

$$Mu_k = z_k Nu_k, \quad Z = \sum u_k u_k'.$$

Since (5.43) with $Q = 0$, $\Delta \rightarrow 0$ yields

$$\text{trace}(ZP) = \sum_{k=1}^p u_k' P u_k \leq 0,$$

at least one of vectors u_k is not equal to zero and satisfies $u_k' P u_k \leq 0$. Since $M u_k = z_k N u_k$, this contradicts the assumption (a).

5.3.2 Checking positivity

This is the standard part of the KYP Lemma used for checking positivity.

Theorem 5.2 *Let $A, B, G = G', L, R = R'$ be real matrices of dimensions n -by- n , n -by- m , n -by- n , n -by- m , and m -by- m , such that $R > 0$ and matrix A has no eigenvalues on the imaginary axis. Then the following conditions are equivalent:*

- (a) *strict frequency domain inequality (5.17) is satisfied for some $\epsilon > 0$ for $\Pi = \Pi(j\omega)$ defined by (5.20);*
- (b) *LMI (5.27) and, equivalently, Riccati inequality (5.29) with the coefficients defined by (5.30), has a solution $Q = Q'$;*
- (c) *Hamiltonian matrix (5.36) has no eigenvalues on the imaginary axis.*

Proof of (b) $\Leftrightarrow$ (c). Applying the Schur complement determinant formula

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \det(a) \det(d - ca^{-1}b) = \det(d) \det(a - bd^{-1}c),$$

valid when $\det(a) \neq 0$ and $\det(d) \neq 0$, to

$$a = \begin{bmatrix} A - j\omega I & 0 \\ G & j\omega I - A' \end{bmatrix}, \quad b = \begin{bmatrix} B \\ L \end{bmatrix}, \quad c = \begin{bmatrix} L' & -B' \end{bmatrix}, \quad d = R,$$

yields

$$\det(R) \det(H - j\omega I) = \det(A - j\omega) \det(-A - j\omega) \det \Pi(j\omega).$$

Hence, if $j\omega$ is an eigenvalue of H then $\Pi(j\omega)$ is not positive definite. Conversely, if H has no eigenvalues on the imaginary axis then $\det \Pi(j\omega) \neq 0$ for all $\omega \in \mathbf{R}$, and hence the continuous Hermitian matrix-valued function $\omega \mapsto \Pi(j\omega)$ does not change sign for $\omega \in \mathbf{R}$. Since $\Pi(j\omega) \rightarrow R$ as $\omega \rightarrow \infty$, this guarantees that $\Pi(j\omega)$ is uniformly positive definite.

Proof of (b) $\Leftrightarrow$ (a). This is a special case of Theorem 5.1, with

$$M = \begin{bmatrix} A - I & B \end{bmatrix}, \quad N = \begin{bmatrix} A + I & B \end{bmatrix}, \quad P = \begin{bmatrix} G & L \\ L' & R \end{bmatrix}.$$

5.3.3 A lemma on Hamiltonian matrices

Before we proceed to existence of stabilizing solution of Riccati equations, we need the following important property of stable invariant subspaces of arbitrary Hamiltonian matrices.

Theorem 5.3 *Let $\alpha = \alpha', \beta, \theta = \theta'$ be real n -by- n matrices. Let $x_i = x_i(t)$, $\psi_i = \psi_i(t)$ for $i = 1, 2$ be solutions of the differential equation*

$$\frac{d}{dt} \begin{bmatrix} x_i(t) \\ \psi_i(t) \end{bmatrix} = H \begin{bmatrix} x_i(t) \\ \psi_i(t) \end{bmatrix}, \quad (5.44)$$

where H be the Hamiltonian matrix (5.36), such that $x_i(t) \rightarrow 0$ and $\psi_i(t) \rightarrow 0$ as $t \rightarrow +\infty$. Then

$$\psi_1(0)'x_2(0) = - \int_0^\infty \{x_1'\alpha x_2 + \psi_1'\theta\psi_2\}dt. \quad (5.45)$$

In particular, Theorem 5.3 implies that

$$\psi_1'x_2 = \psi_2'x_1 \quad (5.46)$$

for every pair of vectors

$$\begin{bmatrix} x_1 \\ \psi_1 \end{bmatrix}, \quad \begin{bmatrix} x_2 \\ \psi_2 \end{bmatrix}$$

in the stable invariant subspace $\mathcal{H}_+$ of H .

Proof. The theorem follows easily from the observation that, subject to (5.44),

$$\frac{d}{dt}\psi_1'x_2 = \psi_1(\beta x_1 + \theta\psi_2) + (\theta x_1 - \beta'\psi_1)'x_2 = \psi_1'\theta\psi_2 + x_1'\alpha x_1.$$

5.3.4 Stabilizing solutions of ARE with stabilizable $\theta \geq 0$

The following statement provides necessary and sufficient conditions of existence of a stabilizing solution of ARE (5.31) in the case when $\theta \geq 0$.

Theorem 5.4 *Let $\alpha = \alpha', \beta, \theta = \theta'$ be real n -by- n matrices. Assume that $\theta \geq 0$ is positive semidefinite. Then the following conditions are equivalent:*

- (a) *ARE (5.31) has a stabilizing solution;*
- (b) *The pair (β, θ) is stabilizable and the Hamiltonian matrix (5.36) has no eigenvalues on the imaginary axis.*

Proof: (a) $\Rightarrow$ (b). According to the arguments given in the discussion of Hamiltonian matrices before, H has no eigenvalues on the imaginary axis. Since $\beta - \theta Q$ is a Hurwitz matrix for a stabilizing solution Q , stabilizability of (β, θ) follows.

Proof: (b) $\Rightarrow$ (a). Let $\mathcal{H}_+$ be the stable invariant subspace of H . Since H has no eigenvalues on the imaginary axis, $\mathcal{H}_+$ has dimension n . Hence one of the following two statements must be true:

- (i) for every $x \in \mathbf{R}^n$ there exists a unique $\psi \in \mathbf{R}^n$ such that $[x; \psi] \in \mathcal{H}_+$;
- (ii) there exists $\psi \in \mathbf{R}^n, \psi \neq 0$ such that $[0; \psi] \in \mathcal{H}_+$.

If (i) holds then the map $x \mapsto \psi$ is linear, i.e. $\psi = Mx$. Then, using Theorem 5.3, $(Mx_1)'x_2 = (Mx_2)'x_1$ for all $x_1, x_2 \in \mathbf{R}^n$, i.e. $M = -Q$ where $Q = Q'$ is a symmetric matrix. Since the columns of matrix $[I; -Q]$ span $\mathcal{H}_+$, we have

$$H \begin{bmatrix} I \\ -Q \end{bmatrix} = \begin{bmatrix} I \\ -Q \end{bmatrix} S$$

for some Hurwitz matrix S . In terms of $\alpha = \alpha', \beta, \theta = \theta'$ this means

$$\beta - \theta Q = S, \quad \alpha + \beta' Q = -QS,$$

and hence $Q = Q'$ is the stabilizing solution of the Riccati equation.

If (ii) holds, consider the subspace $\mathcal{V}$ consisting of all vectors $\psi \in \mathbf{R}^n$ such that $[0; \psi] \in \mathcal{H}_+$. Since $\mathcal{H}_+$ is H -invariant, for an arbitrary element $[0; \psi] \in \mathcal{H}_+$, vector $H[0; \psi] = [\theta\psi; -\beta'\psi]$ is also in $\mathcal{H}_+$, and hence, according to Theorem 5.3, $0'(-\beta'\psi) = (\theta\psi)'\psi$, i.e. $\psi'\theta\psi = 0$. Since, by assumption, $\theta = \theta' \geq 0$, this implies $\theta\psi = 0$, and hence $[0; -\beta'\psi] \in \mathcal{H}_+$. Therefore $\mathcal{V}$ is a non-zero stable invariant subspace of $-\beta'$ which is orthogonal to θ : a contradiction with the assumption of stabilizability of pair (β, θ) .

5.3.5 Stabilizing solutions of the dual ARE

By an ARE which is *dual* to (5.31), we mean the ARE

$$\theta - \beta P - P\beta' = P\alpha P. \quad (5.47)$$

As usually, a solution $P = P'$ of (5.47) is called *stabilizing* when $-\beta' - \alpha P$ is a Hurwitz matrix. There is a simple relation between *non-singular* solutions of primal and dual ARE.

Theorem 5.5 *$Q = Q'$ is a non-singular stabilizing solution of ARE (5.31) if and only if $P = Q^{-1}$ is a non-singular stabilizing solution of ARE (5.47).*

Proof. Multiplying (5.31) by $P = Q^{-1}$ on both sides yields (5.47). Similarly, Multiplying (5.47) by $Q = P^{-1}$ on both sides yields (5.31). Moreover, (5.31) can be re-written as

$$Q(\beta - \theta Q)Q^{-1} = -\beta' - \alpha Q^{-1},$$

hence $-\beta' - \alpha P$ is similar to $\beta - \theta Q$, i.e. $-\beta' - \alpha P$ is a Hurwitz matrix if and only if $\beta - \theta Q$ is. ■

When the pair (β, θ) is not stabilizable, ARE (5.31) cannot have a stabilizing solution. Nevertheless, sufficient conditions for existence of a stabilizing solution of the *dual* ARE (5.47) can be provided.

Theorem 5.6 *Let $\alpha = \alpha', \beta, \theta = \theta'$ be real n -by- n matrices. Assume that $\theta \geq 0$ is positive semidefinite, the pair (β, θ) has no uncontrollable modes on the imaginary axis (i.e. $\beta + \theta k$ has no purely imaginary eigenvalues for some matrix k), and the strict Riccati inequality (5.29) has a positive definite solution $Q = Q' > 0$. Then ARE (5.47) has a stabilizing solution $P = P'$ such that $0 \leq P < Q^{-1}$.*

Proof. According to Theorem 5.2, feasibility of (5.29) implies absence of purely imaginary eigenvalues of the Hamiltonian matrix H . Hence the stable H -invariant subspace $\mathcal{H}_+$ has dimension n .

In addition, for every square integrable $x, \psi \in L_{2e}(\mathbf{R}^n)$ satisfying

$$\dot{x}(t) = \beta x(t) + \theta \psi(t),$$

(5.29) yields

$$\begin{aligned}
x(0)'Qx(0) &= \int_0^\infty \{-2x'Q(\beta x + \theta\psi)\}dt \\
&\leq \int_0^\infty \{x'\alpha x - x'Q\theta Qx - 2x'Q\theta\psi\}dt \\
&\leq \int_0^\infty \{x'\alpha x + \psi'\theta\psi\}dt.
\end{aligned}$$

Therefore, if $x, \psi \in L_{2e}(\mathbf{R}^n)$ are square integrable signals satisfying

$$\frac{d}{dt} \begin{bmatrix} x \\ \psi \end{bmatrix} = H \begin{bmatrix} x \\ \psi \end{bmatrix},$$

the inequality $x(0)'Qx(0) \leq -\psi(0)'x(0)$ holds. In other words,

$$x'Qx \leq -\psi'x \quad \forall \begin{bmatrix} x \\ \psi \end{bmatrix} \in \mathcal{H}_+. \quad (5.48)$$

Since $Q > 0$ is strictly positive definite, (5.48) implies that the only $x \in \mathbf{R}^n$ such that $[x; 0] \in \mathcal{H}_+$ is $x = 0$. Accordingly, the elements of $\mathcal{H}_+$ have parameterization

$$\mathcal{H}_+ = \left\{ \begin{bmatrix} -P\psi \\ \psi \end{bmatrix} : \psi \in \mathbf{R}^n \right\}, \quad (5.49)$$

and, using the technique from the proof of Theorem 5.4, it can be shown that $P = P'$ is a stabilizing solution of ARE (5.47). Finally, combining (5.48) with (5.49) yields

$$\psi'PQP\psi \leq \psi'P\psi \quad \forall \psi \in \mathbf{R}^n,$$

and hence $P \geq 0$ is positive semidefinite. ■

6 H2 optimization

This section presents a derivation of optimal solution in the classical H2 optimization problem. Certain properties of the H2 optimal controller are also discussed.

6.1 Derivation of H2 optimal controller

The path developed below relies on reduction of output feedback design to two abstract H2 optimization problems, solved via two algebraic Riccati equations. The reduction is based on studying *impulse responses* of the closed loop system.

6.1.1 Background: impulse responses

For a strictly proper rational transfer matrix

$$G(s) = c(sI - a)^{-1}b,$$

its *impulse response* is the function $g = g(t)$ defined for $t \geq 0$ by

$$g(t) = ce^{at}b \quad (t \geq 0),$$

where

$$e^{at} = I + at + a^2t^2/2 + \cdots + a^kt^k/k! + \cdots$$

is the *matrix exponent* of at , the solution $E(t) = \exp(at)$ of matrix ODE

$$\dot{E}(t) = aE(t), \quad E(0) = I.$$

In alternative terms, $g(t)$ can be defined by

$$g(t) = c\eta(t), \quad \dot{\eta}(t) = a\eta(t), \quad \eta(0) = b. \quad (6.1)$$

A strictly proper rational transfer matrix $G = G(s)$ defines a stable input/output relation if and only if the corresponding impulse response $g = g(t)$ is square integrable over $(0, \infty)$. Moreover, according to the Parseval identity, H2 norm of a stable G equals the integral of square of the *Frobenius norm*

$$\|M\|_F^2 = \|M'\|^2 = \text{trace}(M'M) = \text{trace}(MM')$$

of g :

$$\|G\|_{H2}^2 \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} \|G(j\omega)\|_F^2 d\omega = \int_0^{\infty} \|g(t)\|_F^2 dt \stackrel{\text{def}}{=} \|g\|_{L2}^2.$$

When G has $m > 1$ inputs, i.e.

$$b = \begin{bmatrix} b_1 & b_2 & \dots & b_m \end{bmatrix},$$

where b_i are column vectors, both equation (6.1) and the expression for the H2 norm of G can be separated according to

$$g = \begin{bmatrix} g_1 & g_2 & \dots & g_m \end{bmatrix}, \quad g_i = c\eta_i, \quad \dot{\eta}_i = a\eta_i, \quad \eta_i(0) = b_i, \quad \|g(t)\|_F^2 = \sum_i |g_i(t)|^2.$$

6.1.2 Closed loop impulse responses

Combining LTI plant equations

$$\dot{x} = Ax + B_1w + B_2u, \quad (6.2)$$

$$e = C_1x + D_{12}u, \quad (6.3)$$

$$y = C_2x + D_{21}w, \quad (6.4)$$

with equations of a finite order strictly proper LTI controller

$$\dot{x}_f = A_fx_f + B_fy, \quad (6.5)$$

$$u = C_fx_c \quad (6.6)$$

yields a state space LTI model with state $x_c = [x; x_f]$. The following statement (a time-domain version of the Q-parameterization theorem) provides a complete description of closed loop impulse responses $X = X(t)$ (from w to x) and $U = U(t)$ (from w to u).

Lemma 6.1 *Assume that controller (6.5), (6.6) stabilizes system (6.2), (6.3), (6.4), i.e.*

$$A_{cl} \stackrel{\text{def}}{=} \begin{bmatrix} A & B_2C_f \\ B_fC_2 & A_f \end{bmatrix}$$

is a Hurwitz matrix. Then

$$\dot{X}(t) = AX(t) + B_2U(t), \quad X(0) = B_1, \quad (6.7)$$

and there exist square integrable (over $(0, \infty)$) matrices $\Psi = \Psi(t)$ and $V = V(t)$ such that

$$\begin{bmatrix} X(t) \\ U(t) \end{bmatrix} = \Psi(t)B_1 + V(t)D_{21}, \quad (6.8)$$

$$\dot{\Psi}(t) = \Psi(t)A + VC_2, \quad \Psi(0) = \begin{bmatrix} I \\ 0 \end{bmatrix}. \quad (6.9)$$

Proof. By the definition,

$$\begin{bmatrix} X(t) \\ U(t) \end{bmatrix} = \exp \left(t \begin{bmatrix} A & B_2 C_f \\ B_f C_2 & A_f \end{bmatrix} \right) \begin{bmatrix} B_1 \\ B_f D_{21} \end{bmatrix}.$$

Then (6.7) follows by the definition of the matrix exponent, and (6.8), (6.9) hold for

$$\Psi(t) = \exp \left(t \begin{bmatrix} A & B_2 C_f \\ B_f C_2 & A_f \end{bmatrix} \right) \begin{bmatrix} I \\ 0 \end{bmatrix}, \quad V(t) = \exp \left(t \begin{bmatrix} A & B_2 C_f \\ B_f C_2 & A_f \end{bmatrix} \right) \begin{bmatrix} 0 \\ B_f \end{bmatrix}.$$

■

6.1.3 Reduction to abstract H2 optimization

H2 norm of the closed loop system is the square $\|E^u\|_{L^2}^2$ of L2 norm of

$$E^u(t) \stackrel{\text{def}}{=} C_1 X(t) + D_{12} U(t).$$

According to Lemma 6.1, the columns e_i^u of E^u can be represented as

$$e_i^u = C_1 x_i + D_{12} u_i, \quad \dot{x}_i = A x_i + B_2 u_i, \quad x_i(0) = b_i,$$

where x_i , u_i , and b_i are the columns of X , U , and B_1 respectively. As long as minimization of L2 norms of e_i^u is concerned, the abstract H2 optimization problem

$$\int_0^\infty |C_1 x + D_{12} u|^2 dt \rightarrow \min, \quad \dot{x} = A x + B_2 u, \quad x(0) = x_0, \quad \lim_{t \rightarrow \infty} x(t) = 0$$

becomes relevant. Assume that D_{12} is left invertible, and a stabilizing solution $P_u = P_u'$ of the corresponding Riccati equation does exist, and hence the “completion of squares”

$$|C_1 x + D_{12} u|^2 + 2x' P_u (A x + B_2 u) = |D_{12}(u - Fx)|^2$$

is possible, where $A + B_2 F$ is a Hurwitz matrix. Then

$$\|E^u\|_{L^2}^2 = \text{trace}(B_1' P_u B_1) + \|D_{12}(U - FX)\|_{L^2}^2. \quad (6.10)$$

Existence of a controller (6.5), (6.6) for which $U(t) \equiv FX(t)$ would imply that the trace of $B_1' P_u B_1$ is the minimum of $\|E^u\|_{L^2}^2$, and the optimal feedback controller is $u(t) = Fx(t)$. However, as $Fx(t)$ is not directly measurable according to the setup, the minimum of $\|E^u\|_{L^2}^2$ is usually strictly larger.

Let

$$E^y(t) = (U(t)' - X(t)'F')D'_{12} = \begin{bmatrix} X(t)' & U(t)' \end{bmatrix} Z \quad \text{for} \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} -F' \\ I \end{bmatrix} D'_{12}.$$

As follows from (6.10), minimization of $\|E^u\|_{L_2}^2$ is equivalent to minimization of $\|E^y\|_{L_2}^2$. According to Lemma 6.1, the columns e_i^y of E^y can be represented as

$$e_i^y = B'_1 \eta_i + D'_{21} \xi_i, \quad \dot{\eta}_i = A' \eta_i + C'_2 \xi_i, \quad \eta_i(0) = h_i,$$

where η_i , ξ_i , and h_i are the columns of $\Psi'Z$, $V'Z$, and $Z_1 = -F'D'_{12}$ respectively. As long as minimization of L2 norms of e_i^y is concerned, the abstract H2 optimization problem

$$\int_0^\infty |B'_1 \eta + D'_{21} \xi|^2 dt \rightarrow \min, \quad \dot{\eta} = A' \eta + C'_2 \xi, \quad \eta(0) = \eta_0, \quad \lim_{t \rightarrow \infty} \eta(t) = 0$$

becomes relevant. Assume that D_{21} is right invertible, and a stabilizing solution $P_y = P'_y$ of the corresponding Riccati equation does exist, and hence the “completion of squares”

$$|B'_1 \eta + D'_{21} \xi|^2 + 2\eta' P_y (A' \eta + C'_2 \xi) = |D'_{21} (\xi - L' \eta)|^2 \quad (6.11)$$

is possible, where $A + LC_2$ is a Hurwitz matrix. Then

$$\|E^y\|_{L_2}^2 = \text{trace}(D_{12} F P_y F' D'_{12}) + \|D'_{21} (V'Z - L' \Psi'Z)\|_{L_2}^2. \quad (6.12)$$

6.1.4 Optimal H2 controller

Representations (6.10) and (6.12) suggest that minimum in H2 optimization is given by

$$\min \|E^u\|_{L_2}^2 = \text{trace}(B'_1 P_u B_1) + \text{trace}(D_{12} F P_y F' D'_{12}),$$

and is achieved by a controller for which $V(t)'Z = L'\Psi(t)'Z$ for all t . By inspection, the stabilizing feedback law

$$u(t) = F\hat{x}(t), \quad (6.13)$$

$$\frac{d\hat{x}(t)}{dt} = A\hat{x}(t) + B_2 u(t) + L(C_2 \hat{x}(t) - y(t)), \quad (6.14)$$

satisfies this condition, and, therefore, is optimal.

Controller (6.13), (6.14) uses a *state observer* (6.14), which, as is easy to show using (6.11), is optimal in the H2 norm sense. Equation (6.13) substitutes the optimal estimate $\hat{x}(t)$ of $x(t)$ into the formula of the H2 optimal full state feedback controller $u = Fx$.

6.2 Properties of H2 optimal controllers

Recall that for the standard setup of minimizing the closed loop H2 norm from w to z in (6.2),(6.3) the optimal LTI controller is given by

$$u = F\hat{x}, \quad (6.15)$$

$$\dot{\hat{x}} = A\hat{x} + B_2u + L(C_2\hat{x} - y), \quad (6.16)$$

where F is such that

$$|C_1x + D_{12}u|^2 + 2x'P_u(Ax + B_2u) = |D_{12}(u - Fx)|^2, \quad (6.17)$$

and $A + B_2F$ is a Hurwitz matrix, where D_{12} is left invertible. Similarly, L is such that

$$|B'_1\eta + D'_{21}\xi|^2 + 2\eta'P_y(A'\eta + C'_2\xi) = |D'_{21}(\xi - L'\eta)|^2, \quad (6.18)$$

where $A + LC_2$ is a Hurwitz matrix, and D_{21} is right invertible.

In this section, we investigate the closed loop properties of H2 optimal feedback systems which can be derived from these characterizations.

6.2.1 Closed Loop Poles

The closed loop system dynamics can be written in terms of the new states $e = x - \hat{x}$ and $\hat{x}$:

$$\dot{e} = (A + LC_2)e + (B_1 + LD_{21})w, \quad (6.19)$$

$$\dot{\hat{x}} = (A + B_2F)\hat{x} - LC_2e - LD_{21}w. \quad (6.20)$$

In particular, this implies that the closed loop poles are eigenvalues of $A + B_2F$ and $A + LC_2$.

It turns out that identities (6.17), (6.18) can be used to get some insight into locations of these poles. Indeed, eigenvalues of $A + B_2K$ are zeros of

$$\psi_u(s) = I - F(sI - A)^{-1}B_2,$$

because

$$\det(sI - A - B_2F) = \det(sI - A) \det(I - B_2F(sI - A)^{-1}) = \det(sI - A)^{-1} \det(\psi_u(s)),$$

where we have use the fact that

$$\det(I + ab) = \det(I + ba)$$

for every pair of matrices a, b of compatible dimensions (and the identity matrices on both sides may have different dimensions). Similarly, eigenvalues of $A + LC_2$ are zeros of

$$\psi_y(s) = I - L'(sI - A')^{-1}C'_2.$$

On the other hand, noting that (6.17) will hold for complex vectors x, u , as long as it is modified to a Hermitian norm identity

$$|C_1x + D_{12}u|^2 + 2\text{Re}\{x'P_u(Ax + B_2u)\} = |D_{12}(u - Fx)|^2, \quad (6.21)$$

where $'$ denotes Hermitian conjugation, and substituting

$$x = (j\omega I - A)^{-1}B_2u$$

yields the identity

$$P_{12}(-s)^T P_{12}(s) = \psi_u(-s)^T D'_{12} D_{12} \psi_u(s) \quad (6.22)$$

for all $s = j\omega$ on the imaginary axis (except, strictly speaking, possible eigenvalues of A). Since (6.22) is an identity between rational functions, its validity on the imaginary axis implies its validity for all s . In particular, *zeros of $\det(\psi_u(s))$ (i.e. the eigenvalues of $A + B_2F$) are the stable zeros of $\det\{P_{12}(-s)^T P_{12}(s)\}$.*

A similar derivation shows that *zeros of $\det(\psi_y(s))$ (i.e. the eigenvalues of $A + LC_2$) are the stable zeros of $\det\{P_{21}(s)P_{21}(-s)^T\}$.*

For example, in a special case when the optimization setup is defined by a SISO plant

$$P_0(s) = C(sI - A)^{-1}B,$$

where (A, B) is controllable and (C, A) is observable, according to

$$\dot{x} = Ax + B(w_1 + u), \quad (6.23)$$

$$z_1 = Cx, \quad (6.24)$$

$$z_2 = r_z u, \quad (6.25)$$

$$y = Cx + r_y w_2, \quad (6.26)$$

where $w = [w_1; w_2]$ is the total disturbance, and r_z, r_y are positive coefficients, the closed loop poles of H2 optimal feedback system are the stable zeros of

$$\phi_u(s) = r_z^2 + P_0(s)P_0(-s)$$

and

$$\phi_y(s) = r_y^2 + P_0(s)P_0(-s).$$

In particular, one can now apply the standard properties of the root locus to predict the asymptotic behavior of the poles as r_z and r_y converge to zero or infinity. For example, when r_z is small, some eigenvalues of $A + B_2K$ are close to the stable zeros of $P_0(s)$ and to the mirror (with respect to the imaginary axis) images of unstable zeros of $P_0(s)$, with the rest of eigenvalues approaching infinity. Similarly, when r_z is large, the eigenvalues of $A + BK_2$ approximate stable eigenvalues of A and the mirror images of unstable eigenvalues of A .

6.2.2 Closed Loop Stability Robustness

Is stability of the closed loop be preserved when the optimal gains K, L are replaced by some other gains? Optimal H2 controllers frequently have plenty of robustness this way, though this usually does not transform into real robustness (which is caused by errors in modeling the *outside* world, not the controller's coefficients).

Consider again the special feedback optimization setup (6.23)-(6.26). The corresponding solutions P_u and P_y of Riccati equations satisfy

$$P_u(A + BF) + (A + BF)'P_u = -C'C - r_z^2 F'F, \quad B'P_u = -r_z^2 F, \quad (6.27)$$

$$P_y(A' + L'C') + (A' + L'C')P_y = -BB' - r_y^2 LL', \quad C_2P_y = -r_y^2 L', \quad (6.28)$$

and hence both are positive definite. Moreover (6.27) implies that

$$P_u(A + BzF) + (A + BzF)'P_u = -C'C - r_z^2 F'(1 + 2\text{Re}(z))F$$

is negative definite for every complex number z with $\text{Re}(z) > -0.5$. Therefore, all eigenvalues of $A + BzF$ have negative real part for $\text{Re}(z) > -0.5$. Similarly, $A + LzC_2$ is a Hurwitz matrix for $\text{Re}(z) > -0.5$.

This observation was originally used to claim that the optimal H2 feedback has a 50 percent gain margin and a 60 percent phase margin. In reality, this only holds for the full information or state estimation problems. Indeed, K enters the formula for the output feedback controller in two places: in defining the control output $u = F\hat{x}$, and in defining the state observer dynamics

$$\dot{\hat{x}} = (A + B_2F + LC_2)\hat{x} - Ly.$$

While an *unknown* controller gain uncertainty can be combined with F in the first appearance, in most situations one cannot do the same in the second appearance. The situation with the observer gain is similar. All the controversy around the originally claimed “robustness” of optimal H2 control has served as a catalyst for developing H-Infinity optimization and mu-synthesis.

7 H Infinity optimization

This section covers the algorithmic side of H-Infinity optimization. It presents classical necessary and sufficient conditions of existence of a γ -suboptimal controller, stated in terms of stabilizing solutions of Riccati equations, and provides explicit formulae for the so-called “central” γ -suboptimal controller.

7.1 Problem Formulation and Algorithm Objectives

This section examines the specifics of the way in which the H-Infinity optimal LTI feedback design problem is formulated (and, eventually, solved).

7.1.1 Suboptimal Feedback Design

H-infinity optimization problem appears to be formulated as the task of designing a stabilizing controller $K(s)$, which minimizes the H-infinity norm of the closed-loop transfer matrix T_{ew} from w to e for a given open loop plant $P(s)$ (see Figure 7.22), defined by state

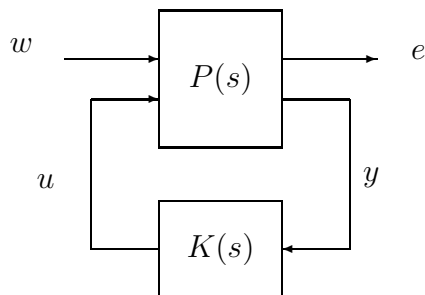


Figure 7.22: General LTI design setup

space equations

$$\begin{aligned}
 \dot{x} &= Ax + B_1w + B_2u \\
 e &= C_1x + D_{11}w + D_{12}u \\
 y &= C_2x + D_{21}w
 \end{aligned}$$

A set of standard well posedness constraints is imposed on the setup:

- for output feedback stabilizability, the pairs (A, B_2) and (C_2, A) must be respectively stabilizable and detectable

- for non-singularity, D_{21} must be right invertible (full measurement noise), D_{12} must be left invertible (full control penalty), and matrices

$$\begin{bmatrix} A - sI & B_2 \\ C_1 & D_{12} \end{bmatrix}, \quad \begin{bmatrix} A - sI & B_1 \\ C_2 & D_{21} \end{bmatrix}$$

must be respectively left and right invertible for all $s \in j\mathbf{R}$

However, in contrast with the case of H2 optimization, basic H-Infinity algorithms solve a *suboptimal* H-Infinity controller design problem, formulated as that of finding whether, for a given $\gamma > 0$, a controller achieving the closed loop L2 gain $\|T_{ew}\|_\infty < \gamma$ exists, and, in case the answer is affirmative, calculating one such controller.

7.1.2 Why Suboptimal Controllers?

There are several reasons to prefer suboptimal controllers over the optimal one in H-Infinity optimization. One of the most compelling reasons is that the optimal closed loop transfer matrix T_{ew} can be shown to have a *constant* largest singular number over the complete frequency range. In particular, this means that the optimal controller is not strictly proper, and the optimal frequency response to the cost output does not roll off at high frequencies.

Example. Consider the standard LTI feedback optimization setup with

$$P(s) = \begin{bmatrix} \frac{1}{\frac{s+1}{1-s}} & -1 \\ \frac{1-s}{1+s} & 0 \end{bmatrix},$$

(a case of state estimation). Since there is no feedback loop to close here, the controller transfer function $K(s)$ must be a stable, and the cost is the H-Infinity norm

$$\|T_{ew}\|_\infty \rightarrow \min, \quad T_{ew} = \frac{1-s}{1+s}K(s) - \frac{1}{1+s}.$$

Since $T_{zw}(1) = 0.5$ for all K , and $\|T\|_\infty \geq |T(1)|$ for all T , the norm $\|T_{zw}\|_\infty$ cannot be less than 0.5. Also, there is the only way to make it identically equal to 0.5, by using $K_0(s) \equiv 0.5$. Hence $K_0(s) = 0.5$ is the (only) H-Infinity optimal controller, and the optimal T_{zw} has a flat (at 0.5) amplitude of frequency response.

Note that the H2 optimal $K(s)$ equals zero in this case.

Another good reason is that, in a setup with more than one control variable or more than one sensor variable, the optimal H-Infinity controller is frequently not unique (indeed,

there could be a continuum of H-Infinity optimal controllers). While some of those optimal controllers could be arguably better than the other, the “most optimal” ones tend to have an order much larger than the necessary minimum (which equals the order of the plant).

Example. Consider the standard LTI feedback optimization setup with 2 sensors, 2 controls, and with

$$P(s) = \begin{bmatrix} \frac{1}{s+1} & 0 & -1 & 0 \\ 0 & \frac{1}{2} \frac{1}{s+1} & 0 & -1 \\ \frac{1-s}{1+s} & 0 & 0 & 0 \\ 0 & \frac{1-s}{1+s} & 0 & 0 \end{bmatrix}$$

(this is essentially the optimization task from the previous example, repeated twice with some scaling). Due to the same arguments as before, the closed loop H-Infinity norm from w to e cannot be made smaller than 0.5. However, there are plenty of “controllers” $K(s)$ which will make the closed loop H-Infinity norm exactly equal to 0.5: for example, one can take

$$K(s) = \begin{bmatrix} 0.5 & 0 \\ 0 & k_2 \end{bmatrix},$$

which will be an optimal controller for every $k_2 \in [0, 0.5]$.

A somewhat less compelling reason for not using H-Infinity optimal controllers is that, while it is known how to find them using ideal algebraic calculations, implementing these formulae in an efficient numerical algorithm is a tough problem.

7.1.3 What is done by the software

The μ -Analysis and Synthesis Toolbox provides function `hinfsyn.m` for calculating a suboptimal output feedback:

```
function [k,g,gfin]=hinfsyn(p,nmeas,ncon,gmin,gmax,tol)
```

(some additional optional inputs and outputs are possible here). Here “p” is the packed model of $P(s)$, “ncon” and “nmeas” are the numbers of actuators and sensors, “gmin” and “gmax” are a lower and an upper a-priori bounds of the achievable closed-loop H-infinity norm. The program performs a binary search for the “level of optimality” parameter γ . At any point, it operates with a “current guess” γ of the minimal achievable closed loop H-infinity norm, a lower bound γ_- and an upper bound γ_+ of γ . The initial values of γ_- and γ_+ are supplied by “gmin” and “gmax”, and, initially, $\gamma = \gamma_+$. At each iteration of the algorithm, it is checked whether it is possible to design a controller with the resulting

closed loop H-infinity gain less than γ . If the answer is positive, new values of γ and γ_+ are assigned, according to

$$\gamma^{new} = 0.5(\gamma_-^{old} + \gamma^{old}), \quad \gamma_+^{new} = \gamma^{old}.$$

If the answer is negative, new values of γ and γ_- are assigned, according to

$$\gamma^{new} = 0.5(\gamma_+^{old} + \gamma^{old}), \quad \gamma_-^{new} = \gamma^{old}.$$

This process is continued until the relative difference between γ_+ and γ_- becomes smaller than the tolerance parameter “tol”. Then, actual suboptimal feedback design is performed, with γ_+ being the target H-infinity performance.

For a given γ , in order to check that a γ -suboptimal solution exists, the algorithm calls for forming two auxiliary abstract H2 optimization problems. Roughly speaking, one of these equations corresponds to the limitations in the “full information feedback” part of the stabilization task, and the other corresponds to the “zero sensor input” part of the stabilization task. For feasibility of the suboptimal H-infinity control problem, stabilizing solutions X and Y of some Riccati equations must exist and be non-negative definite. In addition, a coupling condition $\lambda_{\max}(YX) < \gamma^2$ must be satisfied.

When checking solvability of the Riccati equations, “hinfosyn” uses the “Hamiltonian matrices approach”, which associates stabilizing (i.e. such that $\beta - \theta p$ is a Hurwitz matrix) solutions $p = p'$ of an algebraic Riccati equation

$$\alpha + p\beta + \beta'p = p\theta p$$

with the stable invariant subspace of the auxiliary Hamiltonian matrix

$$H = \begin{bmatrix} \beta & \theta \\ \alpha & -\beta' \end{bmatrix}.$$

Unlike in the Riccati equations and Hamiltonian matrices associated with the usual KYP Lemma, matrix θ will not be positive semidefinite. This will be somewhat compensated by the validity of inequality $\alpha \geq 0$.

7.2 Background Results

This section contains general mathematical results which play an important role in understanding H-Infinity optimization. In addition to providing a specific version of the KYP Lemma and a statement on computing stabilizing solutions of general algebraic Riccati equation, it presents the *Parrot's theorem* and its generalizations.

7.2.1 A special case of the KYP Lemma

As formulated before, KYP Lemma does not provide information about sign definiteness of matrix $p = p'$ of coefficients of the associated quadratic storage function. However, there exists a number of situations when positivity of p is implied by the setup. To derive a solution of H-Infinity optimization problem, we need one such result.

Theorem 7.1 *The following conditions are equivalent:*

- (a) *a is a Hurwitz matrix, and $G(s) = d + c(sI - a)^{-1}b$ has H -Infinity norm less than γ ;*
- (b) *there exists $p = p' > 0$ such that the quadratic form*

$$\sigma_p(x, w) = \gamma^2|w|^2 - |cx + dw|^2 - 2x'p(ax + bw)$$

is strictly positive definite.

Proof. Taking the KYP lemma into account, the only thing to prove here is that, subject to positive definiteness of σ_p , $p > 0$ whenever a is a Hurwitz matrix. Indeed, $\sigma_p \gg 0$ implies the Lyapunov inequality

$$pa + a'p + c'c < 0,$$

hence $p > 0$ if and only if a is a Hurwitz matrix. ■

7.2.2 The Parrot's Lemma and Its Generalizations

The classical Parrot's lemma establishes an explicit solution formulae for the following matrix optimization question: given real matrices a, b, c of dimensions q -by- n , n -by- k , and m -by- n respectively, minimize

$$J(L) = \left\| \begin{bmatrix} a & b \\ c & L \end{bmatrix} \right\|$$

over the set of all m -by- k real matrices L .

Theorem 7.2 *The minimum of $J(L)$ equals*

$$\gamma = \max \left\{ \left\| \begin{bmatrix} a & b \end{bmatrix} \right\|, \left\| \begin{bmatrix} a \\ c \end{bmatrix} \right\| \right\},$$

and is achieved, in particular, with L defined (uniquely) by

$$Ly = \arg \max_v \inf_w \{ \gamma^2(|w|^2 + |y|^2) - |aw + by|^2 - |cw + v|^2 \}.$$

Theorem 7.2 is about defining a “control variable” v as a linear function $v = Lg$ of the “sensor variable” g , with an objective of satisfying a given quadratic inequality $\sigma(f, v, g) \geq 0$, where

$$\sigma(f, v, g) = \gamma^2(|f|^2 + |g|^2) - |af + bg|^2 - |cf + v|^2.$$

Since $g = 0$ would imply $v = Lg = 0$, the inequality $\sigma(f, 0, 0) \geq 0$ must be satisfied, which yields

$$\gamma \geq \left\| \begin{bmatrix} a \\ c \end{bmatrix} \right\|. \quad (7.1)$$

In addition, for every given pair (f, g) , the maximum of $\sigma(f, v, g)$ must be non-negative, which yields the inequality

$$\gamma \geq \left\| \begin{bmatrix} a & b \end{bmatrix} \right\|. \quad (7.2)$$

The non-trivial part of the Parrot’s theorem is concerned with establishing that

$$\sup_v \inf_f \sigma(f, v, g) \geq 0 \quad \forall g.$$

Since (7.2) means that

$$\inf_f \sup_v \sigma(f, v, g) \geq 0 \quad \forall g,$$

the Parrot’s lemma follows from the minimax statement

$$\inf_f \sup_v \sigma(f, v, g) = \sup_v \inf_f \sigma(f, v, g) \quad \forall g,$$

which in turn follows from the standard minimax theorem since $\sigma(f, v, g)$ is strictly concave with respect to v , and (7.1) means that $\sigma(f, v, g)$ is convex with respect to f .

In future derivations, the following generalization of the Parrot’s lemma will be used.

Theorem 7.3 *Let $\sigma = \sigma(f, v, g)$ be a quadratic form which is concave with respect to its second argument (i.e. $\sigma(0, v, 0) \geq 0$ for all v). Then the following conditions are equivalent:*

(a) *there exists matrix L and $\epsilon > 0$ such that*

$$\sigma(f, Lg, g) \geq \epsilon(|f|^2 + |g|^2) \quad \forall f, g;$$

(b) *there exists $\epsilon > 0$ such that*

$$\sigma(f, 0, 0) \geq \epsilon|f|^2, \quad \sup_v \sigma(f, v, g) \geq \epsilon(|f|^2 + |g|^2).$$

Moreover, when (b) is satisfied, L is defined by the condition that the quadratic form

$$\sigma_L(y) = \min_f \sigma(f, Ly, y)$$

is strictly positive definite.

Proof. Following the arguments explaining the classical Parrot's theorem, it is easy to see that (a) implies (b). The fact that (b) implies (a) follows from the standard minimax theorem. ■

7.3 H-Infinity Optimization for a Simplified Setup

In this section, we present a derivation of necessary and sufficient conditions of existence of a γ -suboptimal feedback law, as well as a formula for one such feedback (the so-called *central controller*), for a slightly simplified version of the standard feedback design setup.

7.3.1 The Simplified Setup Solution

To keep the formulae simple, a commonly used *simplified setup* is considered here: noises do not enter the cost, no cross-term penalties between control and state variables are allowed, and sensor noises and plant disturbances are decoupled:

$$\dot{x} = Ax + \bar{B}_1 w_1 + B_2 u, \quad (7.3)$$

$$z = e = \begin{bmatrix} \bar{C}_1 x \\ u \end{bmatrix}, \quad (7.4)$$

$$y = C_2 x + w_2, \quad (7.5)$$

i.e.

$$B_1 = \begin{bmatrix} \bar{B}_1 & 0 \end{bmatrix}, \quad C_1 = \begin{bmatrix} \bar{C}_1 \\ 0 \end{bmatrix}, \quad D_{12} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad D_{21} = \begin{bmatrix} 0 & I \end{bmatrix}.$$

The simplified setup is non-singular if and only if matrices

$$\begin{bmatrix} A - j\omega I & B_1 \end{bmatrix}, \quad \begin{bmatrix} A' - j\omega I & C_1' \end{bmatrix} \quad (7.6)$$

are right invertible for all real ω . In addition, for stabilizing controllers to exist, the pair (A, B_2) must be stabilizable, and the pair (C_2, A) must be detectable.

Consider the following Riccati equations defined by the coefficients of (7.3)-(7.5) and a sub-optimality level $\gamma > 0$:

$$\bar{C}_1' \bar{C}_1 + XA + A'X = X(B_2 B_2' - \gamma^{-2} \bar{B}_1 \bar{B}_1')X, \quad (7.7)$$

$$\bar{B}_1' \bar{B}_1 + Y A' + A Y = Y (C_2 C_2' - \gamma^{-2} \bar{C}_1 \bar{C}_1') Y, \quad (7.8)$$

with respect to symmetric matrices $X = X'$, $Y = Y'$.

Theorem 7.4 *Assume that the pair (A, B_2) is stabilizable, the pair (C_2, A) is detectable, and matrices in (7.6) are right invertible for all $\omega \in \mathbf{R}$. Then an LTI controller $K = K(s)$ such that the feedback $u = Ky$ stabilizes system (7.3)-(7.5) and yields a closed loop L2 gain strictly less than $\gamma > 0$ exists if and only if Riccati equations (7.7), (7.8) have stabilizing solutions $X = X_\infty \geq 0$, $Y = Y_\infty \geq 0$, and*

$$\gamma^2 I > X_\infty^{1/2} Y_\infty X_\infty^{1/2}. \quad (7.9)$$

While the proof of Theorem 7.4 is constructive, and shows a simple way of calculating a suboptimal controller whenever it exists, the following more explicit statement is frequently used.

Theorem 7.5 *Under the conditions for the existence of a suboptimal H-Infinity controller from Theorem 7.4 are satisfied, an LTI controller $u = K(s)y$ is suboptimal if and only if it has a realization*

$$u(t) = -B_2' X_\infty \hat{x}(t) + v(t) \quad (7.10)$$

$$\dot{\hat{x}}(t) = A \hat{x}(t) + \bar{B}_1 \hat{w}_1(t) + B_2 u(t) + L_1 (C_2 \hat{x}(t) - y(t)) + L_2 v(t), \quad (7.11)$$

$$v(t) = \Delta(s)(y(t) - C_2 \hat{x}(t)), \quad (7.12)$$

$$\hat{w}_1(t) = \gamma^{-2} \bar{B}_1' X_\infty \hat{x}(t), \quad (7.13)$$

where

$$L_1 = -(I - \gamma^{-2} Y_\infty X_\infty)^{-1} Y_\infty C_2', \quad L_2 = \gamma^{-2} Y_\infty X_\infty (I - \gamma^{-2} Y_\infty X_\infty)^{-1} B_2,$$

and $\Delta = \Delta(s)$ is a stable LTI system with H-Infinity norm strictly less than γ .

When $\Delta = 0$, Theorem 7.5 yields the so-called *central controller*

The proof of Theorem 7.4, given in the rest of this section, can be extended easily to the general non-simplified setup. In addition, the theorem can be strengthened by noting that using nonlinear controllers does not reduce the minimal achievable L2 gain. Also, a nice parameterization of *all* stabilizin LTI controllers which yield $\|T_{zw}\|_\infty < \gamma$ can be given in terms of X_∞, Y_∞ and the coefficients of (7.3)-(7.5).

7.3.2 Proof of Theorem 7.4

An LTI controller

$$\dot{x}_f = A_f x_f + B_f y, \quad u = C_f x_f + D_f y$$

defines a closed loop system

$$\frac{d\bar{x}}{dt} = a\bar{x} + bw, \quad z = c\bar{x} + dw,$$

with input w , output z , and state $\bar{x} = [x; x_f]$. According to Theorem 7.1, the controller is γ -suboptimal if and only if there exists a matrix $p = p' > 0$ such that the quadratic form $\sigma_p = \sigma_p(f, Lg, g)$ is strictly positive definite for some matrix L , where

$$\sigma_p(f, v, g) = \gamma^2(|w_1|^2 + |y - C_2 x|^2) - |\bar{C}_1 x|^2 - |u|^2 - 2 \begin{bmatrix} x \\ x_f \end{bmatrix}' p \begin{bmatrix} Ax + \bar{B}_1 w_1 + B_2 u \\ h \end{bmatrix},$$

$$f = \begin{bmatrix} w_1 \\ x \end{bmatrix}, \quad v = \begin{bmatrix} h \\ u \end{bmatrix}, \quad g = \begin{bmatrix} x_f \\ y \end{bmatrix},$$

$$L = \begin{bmatrix} A_f & B_f \\ C_f & D_f \end{bmatrix}.$$

According to the generalized Parrot's lemma, this is equivalent to positive definiteness of

$$\sigma_p(f, 0, 0) = \gamma^2(|w_1|^2 + |C_2 x|^2) - |\bar{C}_1 x|^2 - 2x' p_{11}(Ax + \bar{B}_1 w_1)$$

and

$$\sup_v \sigma_p(f, v, g)$$

for some $p = p' > 0$, where p_{11} is the upper left corner of p .

Consider first the condition of positivity of $\sigma_p(f, 0, 0)$. It is equivalent to Riccati inequality

$$\gamma^2 C_2' C_2 - \bar{C}_1' \bar{C}_1 - p_{11} A - A' p_{11} > \gamma^{-2} p_{11} \bar{B}_1 \bar{B}_1' p_{11},$$

or, equivalently,

$$\bar{B}_1 \bar{B}_1' + AY + YA' < Y(C_2' C_2 - \gamma^{-2} \bar{C}_1' \bar{C}_1)Y,$$

where $Y = \gamma^2 p_{11}^{-1} > 0$. According to KYP lemma, existence of such Y is equivalent to existence and positive semi-definiteness of the stabilizing solution Y_∞ of (7.8). Moreover, $Y > Y_\infty$, and the difference $Y - Y_\infty$ can be made arbitrarily small whenever Y_∞ exists.

Now consider the condition of positivity of $\sup_v \sigma_p(f, v, g)$. Note that the supremum equals plus infinity unless

$$p \begin{bmatrix} x \\ x_f \end{bmatrix} = \begin{bmatrix} \psi \\ 0 \end{bmatrix},$$

in which case

$$\sigma_p(f, v, g) = \gamma^2(|w_1|^2 + |y - C_2 q_{11} \psi|^2) - |\bar{C}_1 q_{11} \psi|^2 - |u|^2 - 2\psi'(A q_{11} \psi + \bar{B}_1 w_1 + B_2 u),$$

where q_{11} is the upper left corner of $q = p^{-1}$, and the supremum equals

$$\gamma^2(|w_1|^2 + |y - C_2 q_{11} \psi|^2) - |\bar{C}_1 q_{11} \psi|^2 - 2\psi'(A q_{11} \psi + \bar{B}_1 w_1) + |B_2' \psi|^2.$$

Hence, positive definiteness of the supremum is equivalent to the matrix inequality

$$q_{11} \bar{C}_1' \bar{C}_1 q_{11} + A q_{11} + q_{11} A' < B_2 B_2' - \gamma^{-2} \bar{B}_1 \bar{B}_1',$$

which can be written in an equivalent form

$$\bar{C}_1' \bar{C}_1 + X A + A' X < X (B_2 B_2' - \gamma^{-2} \bar{B}_1 \bar{B}_1') X,$$

where $X = q_{11}^{-1}$. According to KYP lemma, existence of such X is equivalent to existence and positive semi-definiteness of the stabilizing solution X_∞ of (7.7). Moreover, $X > X_\infty$, and the difference $X - X_\infty$ can be made arbitrarily small whenever X_∞ exists.

A direct calculation of a matrix inverse shows that

$$q_{11}^{-1} = p_{11} - p_{12} p_{22}^{-1} p_{12}',$$

where

$$p = \begin{bmatrix} p_{11} & p_{12} \\ p_{12}' & p_{22} \end{bmatrix} > 0.$$

Hence

$$\gamma^2 Y^{-1} = p_{11} > q_{11}^{-1} = X \geq X_\infty.$$

Equivalently,

$$\gamma^2 I > X_\infty^{1/2} Y X_\infty^{1/2},$$

which implies (7.9). This proves necessity of the conditions for existence of a γ -suboptimal controller.

Conversely, when (7.9) is satisfied, inequality $\gamma^2 Y^{-1} > X$ will hold when the solutions X and Y of the corresponding Riccati inequalities are sufficiently close to X_∞ and Y_∞ . Then

$$p = \begin{bmatrix} \gamma^2 Y^{-1} & X - \gamma^2 Y^{-1} \\ X - \gamma^2 Y^{-1} & \gamma^2 Y^{-1} - X \end{bmatrix}$$

is positive definite and the corresponding σ_p satisfies all conditions of the generalized Parrot's lemma, which implies existence of a γ -suboptimal controller.

8 Robustness analysis with quadratic constraints

The idea of *integral quadratic constraints* (IQC) generalizes the notion of L2 power gain, and provides a simple but powerful approach to analysis of those nonlinear, time-varying and uncertain systems for which good finite order LTI approximation is available.

8.1 General principles of IQC analysis

This section describes basic elements of system analysis via integral quadratic constraints.

8.1.1 Definition of IQC

Let $\mathcal{S} = \{z\} \subset L_{2e}(\mathbf{R}^k)$ be a behavioral system model, i.e. simply a set of k -dimensional signals. Let $\sigma : \mathbf{R}^k \mapsto \mathbf{R}$ be a quadratic form. $\mathcal{S}$ is said to *satisfy the IQC defined by σ* if

$$\inf_{T>0} \int_0^T \sigma(z(t))dt > -\infty \quad (8.1)$$

for all $z(\cdot) \in \mathcal{S}$.

In particular, if $\Delta = \{(f(\cdot), y(\cdot))\} \subset L_{2e}(\mathbf{R}^{n_f}) \times L_{2e}(\mathbf{R}^{n_y})$ is a system model with “input” $f \in L_{2e}(\mathbf{R}^{n_f})$ and “output” $y \in L_{2e}(\mathbf{R}^{n_y})$ then declaring that the L2 power gain of Δ is strictly less than γ is equivalent to stating that Δ satisfies the IQC defined by

$$\sigma(\bar{f}, \bar{y}) = \gamma^2 |\bar{f}|^2 - |\bar{y}|^2 \quad (\bar{f} \in \mathbf{R}^{n_f}, \bar{y} \in \mathbf{R}^{n_y}).$$

Thus, integral quadratic constraints generalize the notion of L2 power gains.

8.1.2 IQC for the pure integrator

A very important IQC is a consequence of integration/differentiation relation for a pair of vector signals.

Theorem 8.1 *Let $\mathcal{S} = \{z\} \subset L_{2e}(\mathbf{R}^k)$ be a behavioral system model. Let L and F be two real n -by- k matrices. Assume that for every $z \in \mathcal{S}$ the signals $x = Lz$, $v = Fz$ are such that*

$$x(t_2) - x(t_1) = \int_{t_1}^{t_2} v(t)dt \quad \forall t_1, t_2 \in \mathbf{R}, \quad (8.2)$$

i.e., lightly speaking, $v = dx/dt$. Then, for every n -by- n symmetric matrix $P = P' \geq 0$, the quadraic form

$$\sigma_P(\bar{z}) = 2(L\bar{z})'P(F\bar{z}) \quad (8.3)$$

defines a valid IQC for $\mathcal{S}$.

Proof.

$$\int_0^T \sigma(z(t))dt = \int_0^T 2x(t)'P\dot{x}(t)dt = x(T)'Px(T) - x(0)'Px(0) \geq -x(0)'Px(0) > -\infty.$$

■

8.1.3 System analysis via IQC and convex optimization

The main benefit of using the IQC interpretation of L2 gain conditions is the possibility of producing new IQC by forming convex combinations of known IQC, justified by the following simple observation.

Theorem 8.2 *Let $\mathcal{S} = \{z\} \subset L_{2e}(\mathbf{R}^k)$ be a behavioral system model satisfying the IQC defined by quadratic forms $\sigma_i : \mathbf{R}^k \mapsto \mathbf{R}$ for $k = 1, \dots, N$. Let $\sigma_0 : \mathbf{R}^k \mapsto \mathbf{R}$ be another quadratic form for which there exist constants $d_i \geq 0$ such that*

$$\sigma(\bar{z}) \geq \sum_{i=1}^N d_i \sigma_i(\bar{z}) \quad \forall \bar{z} \in \mathbf{R}^k. \quad (8.4)$$

Then $\mathcal{S}$ satisfies the IQC defined by σ_0 .

Proof. By assumption, for every $z \in \mathcal{S}$ there exist $r_i = r_i(z) \in \mathbf{R}$, $i = 1, \dots, N$, such that

$$\int_0^T \sigma_i \sigma_i(z(t))dt \geq r_i \quad \forall T > 0, i \in \{1, \dots, N\}.$$

Multiplying these inequalities by d_i and adding them together yields

$$\int_0^T \left\{ \sum_{i=1}^N d_i \sigma_i(z(t)) \right\} dt \geq r = \sum_{i=1}^N d_i r_i = r.$$

Hence, because of (8.4),

$$\int_{T>0} \int_0^T \sigma_0(z(t))dt \geq r > -\infty.$$

■

When σ_i for $i = 0, 1, \dots, N$ are given, the search for coefficients $d_i \geq 0$ satisfying (8.4) becomes a semidefinite program. Thus, analysis of a system $\mathcal{S} = \{z(\cdot)\} \subset L_{2e}(\mathbf{R}^k)$ via IQC is performed in three steps.

- Step 1: describe the analysis objective (the statement to be verified) in the form of a quadratic constraint defined by a quadratic form $\sigma_0 : \mathbf{R}^k \mapsto \mathbf{R}$.
- Step 2: provide a sufficiently rich set of quadratic forms $\sigma_i : \mathbf{R}^k \mapsto \mathbf{R}$ defining valid quadratic constraints for $\mathcal{S}$.
- Step 3: use LMI optimization to try to find coefficients $d_i \geq 0$ such that condition (8.4) is satisfied.

If the semidefinite program in step 3 is infeasible, the IQC analysis fails, either because the set of IQC used is not good enough, or simply because the conclusion to be reached is not true. A feasibility conclusion means that a positive statement about the system is proven.

It is common to have extra parameters in the LMI defined by (8.4). For example, in order to derive an L2 power gain bound from input $f(t) = Lz(t)$ to output $y(t) = Cz(t)$ of system $\mathcal{S} = \{z(\cdot)\}$, it is natural to define σ_0 as

$$\sigma_0(\bar{z}) = r|L\bar{z}|^2 - |C\bar{z}|^2,$$

where parameter r represents the square of an L2 power gain bound. Then r will enter linearly into (8.4) and hence can be minimized during LMI optimization with respect to $(r, d_1, \dots, d_N)$ to get a better L2 gain bound $\gamma = r^{1/2}$.

Similarly, if signals $x = Lz$, $v = Fz$ satisfy $dx/dt = v$ for all $z \in \mathcal{S}$, one can use $\sigma_i(\bar{z}) = \bar{z}' L' P F \bar{z}$, where $P = P' \geq 0$ is an arbitrary positive semidefinite matrix. Then, without loss of generality, one can fix the value of d_i at $d_i = 1$, and include $P = P' \geq 0$ among the decision parameters of LMI optimization.

8.1.4 Example: small gain theorem via IQC manipulations

Let $S_0 = \{([f; w], [y; v])\} \subset L_{2e}(\mathbf{R}^{n_f+n_w}) \times L_{2e}(\mathbf{R}^{n_y+n_v})$ be a behavioral system model with input $[f; w]$ and output $[y; v]$, where $f \in L_{2e}(\mathbf{R}^{n_f})$, $w \in L_{2e}(\mathbf{R}^{n_w})$, $y \in L_{2e}(\mathbf{R}^{n_y})$, and $v \in L_{2e}(\mathbf{R}^{n_v})$. Let $\Delta = \{(v, w)\} \subset L_{2e}(\mathbf{R}^{n_v}) \times L_{2e}(\mathbf{R}^{n_w})$ be a behavioral system model with input $v \in L_{2e}(\mathbf{R}^{n_v})$ and output $w \in L_{2e}(\mathbf{R}^{n_w})$. Finally, let behavioral model $S = \{(f, y)\} \subset L_{2e}(\mathbf{R}^{n_f}) \times L_{2e}(\mathbf{R}^{n_y})$ with input $f \in L_{2e}(\mathbf{R}^{n_f})$ and output $y \in L_{2e}(\mathbf{R}^{n_y})$ be the *interconnection* of S_0 and Δ , defined as the set of all pairs (f, y) for which there exist $(v, w) \in \Delta$ such that $([f; w], [y; v]) \in S_0$, as shown on Figure 8.23.

One way to state the classical small gain theorem for this situation is as follows: *if L2 power gains of Δ and Σ_0 are less than 1 then L2 power gain of S is not larger than*

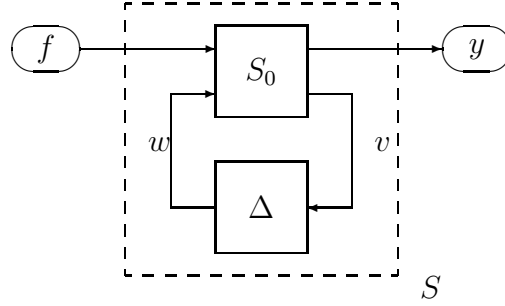


Figure 8.23: Interconnection of the classical small gain theorem

1. IQC manipulations offer a very straightforward proof. Indeed, the assumptions can be stated as two IQC, defined by

$$\sigma_1(f, w, y, v) = |f|^2 + |w|^2 - |y|^2 - |v|^2, \quad \sigma_2(f, w, y, v) = |v|^2 - |w|^2.$$

Summation of these IQC yields a new valid IQC defined by

$$\sigma_0(f, w, y, v) = |f|^2 - |y|^2,$$

which completes the proof.

8.1.5 Example: L2 gain bound with multiple uncertain subsystems

Consider a system model which is defined by the nominal LTI part G described by state space equations

$$\dot{x} = Ax + Bw, \quad v_i(t) = C_i x + D_i w \quad (i = 0, 1, \dots, N), \quad w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{bmatrix}, \quad (8.5)$$

and the structured unmodeled feedback dynamics

$$w_i(\cdot) = \Delta_i(v_i(\cdot)), \quad (8.6)$$

where the only information available about the subsystems Δ_i is that their L2 power gains are smaller than 1. To compute a bound $\gamma = r^{1/2}$ on the L2 power gain from w_0 to v_0 consider the available IQC. The description of Δ_i implies the IQC with

$$\sigma_i(x, w) = |C_i x + D_i w|^2 - |w_i|^2 \quad (i = 1, \dots, N).$$

In addition, the relation between x and $dx/dt = Ax + Bw$ provides the IQC with

$$\sigma_P(x, w) = 2x'P(Ax + Bw) \quad (P = P' \geq 0).$$

The analysis objective is to prove the IQC defined by

$$\sigma_0(x, w) = r|w_0|^2 - |C_0x + D_0w|^2$$

with r being as small as possible. Using convex combinations of IQC yields existence of $d_i \geq 0$ and $P = P' \geq 0$ such that

$$r|w_0|^2 - |C_0x + D_0w|^2 - 2x'P(Ax + Bw) - \sum_{i=1}^N d_i\{|C_i x + D_i w|^2 - |w_i|^2\} \geq 0 \quad (8.7)$$

for all x, w as a certificate of L2 gain from w_0 to v_0 being not larger than $r^{1/2}$. Since (8.7) is equivalent to a matrix inequality which is linear with respect to r, d_i, P , the minimal r can be found by means of LMI optimization.

8.1.6 Dynamic extensions

A key element of IQC analysis is *dynamic extension*, which can be motivated by the following example. Let $\Delta = \{(v, w)\} \subset L_{2e}(\mathbf{R}^1) \times L_{2e}(\mathbf{R}^1)$ be the system with scalar input v and scalar output w , defined by $w(t) = v(t)^3$. An attempt to find quadratic forms $\sigma : \mathbf{R} \times \mathbf{R} \mapsto \mathbf{R}$ which define valid IQC for Δ soon leads to a realization that all such σ satisfy the condition

$$\sigma(\bar{v}, \bar{w}) \geq d\bar{w}\bar{v}$$

for some $d \geq 0$. In other words, the only non-trivial IQC of this form is defined by $\sigma_*(v, w) = wv$.

However, introducing another variable x , defined by

$$\dot{x} = -x + v,$$

changes the situation. Since $(v - x)(v^3 - x^3) \geq 0$ for all v, x ,

$$\frac{dx^4}{dt} = 4(v - x)x^3 \leq 4(v - x)v^3 = 4(v - x)w.$$

Therefore

$$\int_0^T \{4w(v - x)\}dt \geq x(T)^4 - x(0)^4 \geq -x(0)^4 > -\infty,$$

and hence the *dynamically extended* system $\Delta_+ = \{(v, w, x)\}$ satisfies the IQC defined by

$$\sigma_+(v, w, x) = (v - x)w.$$

It can be shown that, after a minor modification, the derivation of the new IQC remains valid when the cubic nonlinearity $w(t) = v(t)^3$ is replaced by an arbitrary monotonic memoryless nonlinearity $w(t) = \phi(v(t))$, where $\phi : \mathbf{R} \mapsto \mathbf{R}$ is a monotonic function such that $\phi(0) = 0$.

In general, adding new signals to a system model by applying stable LTI transformations to existing signals is a major way of enriching the set of available IQC.

8.2 Quadratic constraints for LTI models

For LTI systems, stability and performance analysis can be applied pointwise in the frequency domain. This leads to a non-dynamical interpretation of IQC analysis.

8.2.1 Zero Exclusion Principle

A finite order LTI system is stable if it does not have poles in the closed right half plane

$$\mathbf{C}_+ = \{s \in \mathbf{C} : \operatorname{Re}(s) \geq 0\}.$$

However, checking the *whole* $\mathbf{C}_+$ for the absence of poles is usually inconvenient, especially when the system (and its poles as well) is uncertain. Therefore it is very important to establish that analysis of stability of uncertain LTI systems can be based on excluding marginal instability.

Theorem 8.3 *Let $G = G(s)$, $\Delta = \Delta(s)$ be LTI systems with finite L_2 gains. Assume that*

$$|\det(I - \tau G(j\omega)\Delta(j\omega))| \geq \epsilon > 0 \quad \text{for all } \omega \in \mathbf{R}, \tau \in [0, 1].$$

Then the feedback interconnection of G and Δ has finite L_2 gain.

Theorem 8.3 implies the following version of the *zero exclusion principle*. Let G be a stable finite order LTI system with m inputs and n outputs. Let $\mathbf{\Delta}$ be a *cone* of complex m -by- n complex matrices, i.e. a closed set such that

$$\bar{\Delta} \in \mathbf{\Delta}, \lambda \geq 0 \text{ implies } \lambda \bar{\Delta} \in \mathbf{\Delta}.$$

The interconnection of G and Δ is stable for any transfer function Δ satisfying the condition $\Delta(j\omega) \in \mathbf{\Delta}$ if and only if

$$\det(I - G(j\omega)\bar{\Delta}) \neq 0 \quad \forall \omega \in \mathbf{R} \cup \{\infty\}, \bar{\Delta} \in \mathbf{\Delta}.$$

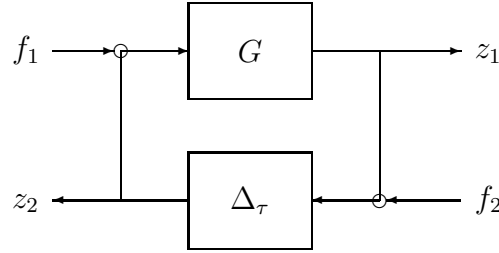


Figure 8.24: a homotopy argument in stability analysis

8.2.2 Structured Singular Values

Let Δ be a cone of complex n -by- m matrices. The *structured singular value*

$$\mu = \mu(M) = \mu(M, \Delta) = \mu_{\Delta}(M)$$

is defined for any complex m -by- n matrix as the reciprocal of the smallest norm of $\Delta \in \Delta$ such that $I - M\Delta$ is *not invertible*. Here Δ defines the *structure* of an uncertain block Δ such that $\Delta(j\omega) \in \mathbf{D}$: the smaller Δ is, the more structure, and the smaller $\mu(M, \Delta)$ is going to be.

Note that, when Δ is the set of *all* n -by- m matrices (i.e. when there is “no structure”), we have

$$\mu(M, \Delta) = \|M\| = \sigma_{\max}(M)$$

which explains the term “structured singular value”

The so-called “complex structured singular value” $\mu_{\mathbf{C}}(M)$ corresponds to the case when

$$\Delta = \left\{ \begin{bmatrix} \Delta_1 & 0 & & 0 \\ 0 & \Delta_2 & & \\ & & \ddots & \\ 0 & & & \Delta_n \end{bmatrix} : \Delta_i \in \mathbf{C} \right\}$$

is the set of all complex diagonal matrices.

The so-called “real structured singular value” $\mu_{\mathbf{R}}(M)$ corresponds to the case when

$$\Delta = \left\{ \begin{bmatrix} \delta_1 & 0 & & 0 \\ 0 & \delta_2 & & \\ & & \ddots & \\ 0 & & & \delta_n \end{bmatrix} : \delta_i \in \mathbf{R} \right\}$$

is the set of all real diagonal matrices.

M. Safonov uses notation $K_m(M, \mathbf{\Delta}) = 1/\mu(M, \mathbf{\Delta})$.

8.2.3 Examples

For

$$M = \begin{bmatrix} j & 100 \\ 0 & j \end{bmatrix}$$

we have

$$\sigma_{max}(M) \approx 100, \quad \mu_{\mathbf{C}}(M) = 1, \quad \mu_{\mathbf{R}}(M) = 0$$

If $\mathbf{\Delta} = \{\Delta = \delta I : \delta \in \mathbf{C}\}$, then $\mu(M, \mathbf{\Delta})$ is the spectral radius of M (i.e. the largest absolute value of an eigenvalue of M).

$$\mu_{\mathbf{R}} \left(\begin{bmatrix} 1 & t \\ -t & 1 \end{bmatrix} \right) = \begin{cases} 0, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

In particular, this demonstrates that some versions of μ are discontinuous functions of its argument. However, $\mu_{\mathbf{C}}(M)$ can be proven to be a continuous function of M .

8.2.4 A “Small μ Theorem”

Consider a feedback interconnection of a given LTI system $G(s)$ and an uncertain LTI system $\Delta(s)$, where $\Delta(s)$ is any *stable* LTI system such that $\Delta(j\omega) \in \mathbf{\Delta}$ for all ω , and $\|\Delta(s)\|_{\infty} < 1$. Then, by the definition of $\mu(G, \mathbf{\Delta})$, and by the “zero exclusion principle”, the interconnection is stable for any expected $\Delta(s)$ if

$$\sup_{\omega} \mu_{\mathbf{\Delta}}(\bar{G}) < 1$$

for any $\bar{G}$ which can be approximated arbitrarily well by the values of $G(j\omega)$.

For example, in the case of “structured” unmodeled dynamics, shown on Figure 8.25, the interconnection is stable and has worst case w_0 to z_0 gain less than one if and only if

$$\sup_{\omega} \mu_{\mathbf{C}}(G(j\omega)) < 1.$$

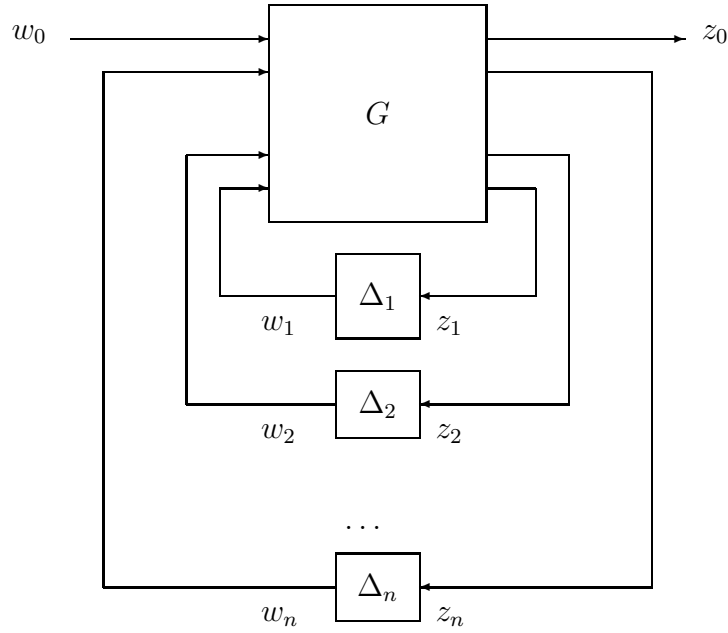


Figure 8.25: small complex mu theorem

8.3 Computation of robustness margins

In general, μ is difficult to compute exactly (such problems are called “NP hard” by computer scientists). In practice, computable upper and lower bounds of μ are used. When those bounds are far apart, a “branch and bound” technique is used.

An *upper bound* of μ is a function $\hat{\mu}$ such that $\hat{\mu}(M) \geq \mu(M)$ for any M . Upper bounds of μ give *sufficient* conditions of stability and robust performance. Lower bounds of μ prove that certain systems are not robustly stable. Usually, calculation of a large lower bound of μ comes with an example of a destabilizing uncertainty.

8.3.1 Lower Bounds of μ

Standard lower bounds for μ are obtained by finding a local minimum in the non-convex optimization problem

$$\rho(\Delta M) \rightarrow \max_{\Delta \in \mathbf{\Delta}}$$

where $\rho(A)$ is the spectral radius of A (when $\mathbf{\Delta}$ is invariant with respect to multiplication by complex scalars) or the real spectral radius of A (for the “real” versions of μ). Roughly

speaking, the search for low bounds of μ amounts to simulations of the uncertain systems with different values of the uncertainty parameters.

For most common structures Δ , the search over Δ can be replaced by the search over the *unitary* elements in Δ , i.e. such that $\Delta\Delta' = I$.

For example, for a 3-by-3 matrix M , figuring out whether $\mu_{\mathbf{R}}(M) < 1$ is equivalent to checking that $I - M\Delta$ is invertible for any diagonal matrix with elements $\delta_i \in [-1, 1]$ on the diagonal ($i = 1, 2, 3$).

Consider separately the 2 cases: $\delta_1 \geq 0$ and $\delta_1 \leq 0$. Since

$$[0, 1] = 0.5 + 0.5 \cdot [-1, 1],$$

$$[-1, 0] = -0.5 + 0.5 \cdot [-1, 1],$$

in each case, the problem can be reduced to checking that $\mu_{\mathbf{R}}(M_{\pm}) < 1$, where

$$M_{\pm} = (I \pm 0.5Me_1e_1')^{-1}M(I - 0.5e_1e_1')$$

e_1 is the first basis vector, “+” corresponds to $\delta_1 \in [0, 1]$, “−” corresponds to $\delta_1 \in [-1, 0]$. For each of $M_{\pm}$, we expect that the gap between upper and lower bounds of μ will be smaller, since the actual range of δ_i has been reduced.

8.3.2 Quadratic Constraints

Matrix $I - M\Delta$ is not invertible if and only if the system of equations

$$y = Mw, \quad w = \Delta y$$

has a non-zero solution (y, w) . To show that this is impossible for $\Delta \in \Delta$, $r\|\Delta\| \leq 1$, we start with finding quadratic forms $\sigma = \sigma(y, w)$ such that

$$\sigma(y, \Delta y) \geq 0 \quad \forall \Delta \in \Delta, \quad \|\Delta\| \leq 1$$

Such conditions are called *quadratic constraints* describing the relation $w = \Delta y$. The idea of “describing” uncertainty using quadratic inequalities is the background of most robustness criteria.

Let Σ be a set of quadratic forms $\sigma(y, w)$ such that

$$\sigma(y, \Delta y) \geq 0 \quad \forall \Delta \in \Delta, \quad \|\Delta\| \leq 1$$

$$\sigma(0, w) < 0 \quad \text{for } w \neq 0$$

Note that any convex combination of such quadratic forms satisfies the condition as well.

Define a functional on Σ by

$$J(\sigma) = \inf\{r \geq 0 : \sigma(Mw, rw) < 0 \quad \forall w \neq 0\}$$

- Functional J is quasi-convex.
- The infimum of J over Σ is an upper bound of $\mu(M, \Delta)$.
- The larger Σ , the better the upper bound.

This is the idea behind upper bounds of μ .

8.3.3 Elementary Uncertainty

Here we derive quadratic constraints for elementary components of uncertainty structures

Unmodeled Dynamics If $w = \Delta y$, where Δ is an arbitrary complex matrix with $\|\Delta\| \leq 1$, the relation between w and y is described by

$$\sigma(y, w) = \|y\|^2 - \|w\|^2 \geq 0$$

Repeated Real Scalar If $w = \delta y$, where $\delta \in [-1, 1]$ is a real number, the relation between w and y is described by

$$\sigma(y, w) = y'Dy - w'Dw + 2\text{Re}(y'Sw) \geq 0$$

where where D, S are arbitrary matrices such that $D = D' \geq 0$ and $S = -S'$.

Quadratic constraints for other elementary uncertainty relations are obtained in a similar way.

8.3.4 An Upper Bound for μ

Quadratic constraints for a general structured uncertainty are obtained as convex combinations of the “elementary” constraints.

Let $\Delta = \{\Delta\}$ be the uncertainty structure for which $w = \Delta y$ means

$$y = [y_1^r; \dots; y_N^r; y_1^c; \dots; y_M^c]$$

$$w = [w_1^r; \dots; w_N^r; w_1^c; \dots; w_M^c]$$

where

$$\begin{aligned} w_k^r &= \delta_k y_k^r, \quad y_k^r, w_k^r \in \mathbf{C}^{n(k)}, \quad \delta_k \in \mathbf{R} \\ w_k^c &= \Delta_k y_k^c, \quad y_k^c \in \mathbf{C}^{m(k)}, \quad w_k^c \in \mathbf{C}^{p(k)} \end{aligned}$$

An upper bound of $\mu(M, \Delta)$ is given by the minimal $r > 0$ such that

$$M'DM + MS + S'M < r^2 \hat{D}$$

where

$$S = \begin{bmatrix} \hat{S} & 0 \\ 0 & 0 \end{bmatrix},$$

$$D = \text{diag}\{D_1, \dots, D_N, d_1 I_{m(k)}, \dots, d_M I_{m(M)}\}$$

$$\hat{D} = \text{diag}\{D_1, \dots, D_N, d_1 I_{p(k)}, \dots, d_M I_{p(M)}\}$$

$$\hat{S} = \text{diag}\{S_1, \dots, S_M\}$$

matrices $D_k = D'_k \geq 0$ and $S_k = -S'_k$ have size $n(k)$ -by- $n(k)$. Optimization of $r_{\min}$ over $D, \hat{D}, S$ is quasi-convex.

The upper bound above was obtained using the quadratic constraint

$$\sigma(y, \Delta y) \geq 0 \quad \forall \Delta \in \mathbf{\Delta}, \quad \|\Delta\| \leq 1$$

where

$$\begin{aligned} \sigma(y, w) &= \sum_{k=1}^N y_k^{r'} D_k y_k^r - w_k^{r'} D_k w_k^r \\ &\quad + 2\text{Re}(y_k^{r'} S_k w_k^r) \\ &\quad + \sum_{k=1}^M d_i (\|y_k^c\|^2 - \|w_k^c\|^2) \end{aligned}$$

8.4 Numerical calculations and examples

Practical verification of *strict* feasibility of IQC models can be done by standard routines of convex optimization. Furthermore, the IQC's known for a number of standard nonlinear, uncertain, or time-varying components, such as nonlinearities with conic, slope, and curvature bounds, LTI uncertainty (real parametric, unmodeled, repeated, uncertain delay), time-varying uncertain gain (fast and slowly time-varying, harmonic, periodic), and many others, can be pre-stored in a computer for easy access. In this way, rather complicated IQC models can be built and tested for feasibility in a very simple manner. On Athena, the software is located in the locker `/mit/6.245`, in the subdirectory `iqc_beta`.

8.4.1 Example: a servo with friction and uncertain delay

Consider a simple model $\mathcal{S}_{\text{servo}}$ of a servo with friction on Figure 8.26, where

$$K(s) = 10 \frac{2s^2 + 2s + 1}{0.01s^2 + s + 0.01} \approx 20s + 20 + 10/s$$

represents a PID controller, the saturated high gain feedback models friction, and $\tau \in [0, 0.05]$ is the uncertain measurement delay. In terms of signals, u is control input, f is

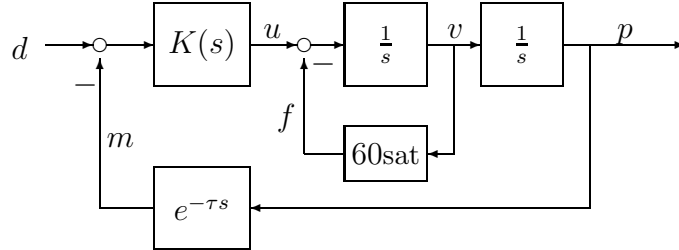


Figure 8.26: Example: a servo with friction

the friction force, v and p are velocity and position while m is a position measurement with uncertain delay. The external input d represents a high-pass sensor noise with the cut-off at 100 rad/sec. The objective is to prove stability of the system and to estimate the noise amplification coefficient J , defined as the maximal L_2 gain in the closed loop channel $e \mapsto p$, where $\dot{d} + 100d = \dot{e}$.

A feasibility test for the IQC model can be specified in the toolbox `iqc_beta` with the commands¹

```
s=tf([1 0],1); % convenient constant
abst_init_iqc; % initialize data structure for handling IQC
e=signal; % external unmodeled disturbance
f=signal; % a nonlinear gain output
r=signal; % an uncertain block output
p=signal; % position (feedback loop signal)
m=p+r; % delayed position
u=(10*(2*s^2+2*s+1)/(0.01*s^2+s+0.01))*((s/(s+100))*e-m);
v=(1/s)*(u-f); % velocity
p==(1/s)*v; % closing the loop
f==60*iqc_sector(v); % f(f-60u)>=0
r==iqc_ltvnorm(p)-p; % ||r+p||^2<=||p||^2
J=iqc_gain_tbx(e,p) % estimate L2 gain e -> p
```

Here the command `abst_init_iqc` initializes the data structure, and `==` is used to close feedback loops. Expression `iqc_sector(v)` produces a signal z_1 that satisfies the IQC

¹This code will only work with the MATLAB version 5.3 or (possibly) higher

defined by the quadratic form $z_1(z_1 - v)$. The essential part of the code of `iqc_sector` can be written as

```
function w=iqc_sector(v)
w=signal; % define the output
c=symmetric; % introduce a scaling constant
c>0; % put an LMI constraint on c
w'*c*(w-v)<0; % describe an IQC constraint
```

Similarly, the expression `iqc_ltvnorm(p)` produces a signal z_2 that satisfies the IQC defined by the quadratic form $|z_2|^2 - |p|^2$. Finally, the command `J=iqc_gain_tbx(e,p)` indicates that $f = e$ is the external disturbance, and $y = p$ is the transient response to be considered.

Running the script produces an empty J , which means that no convex combination of used IQC produces a finite L2 gain bound. This does not necessarily imply that the original servo system is unstable – just that the IQC model is not good enough.

A more accurate IQC model of the servo system would use additional IQC. To check feasibility of the upgraded IQC model, the MATLAB code above should be modified by replacing

```
iqc_sector(v) ... iqc_ltvnorm(p)-p
```

with

```
iqc_monotonic(v) ... iqc\_cdelay(p,.05)
```

respectively. Running the modified script produces $J \approx 58$, which means that the servo system is stable, and the noise amplification coefficient J does not exceed 60.

8.4.2 Example with cubic nonlinearity and delay

For an application of IQC analysis where strict feasibility does not take place consider the following system of differential equations² with an uncertain constant delay parameter τ :

$$\dot{x}_1(t) = -x_1(t)^3 - x_2(t - \tau)^3 \quad (8.8)$$

$$\dot{x}_2(t) = x_1(t) - x_2(t) \quad (8.9)$$

Analysis of this system is easy when $\tau = 0$, and becomes more difficult when τ is an arbitrary constant in the interval $[0, \tau_0]$. The system is not exponentially stable for any

²Suggested by Petar Kokotovich

value of τ . Our objective is to show that, despite the absence of exponential stability, the method of IQC analysis can be applied.

For $\tau = 0$, we begin with describing (8.8),(8.9) by the behavior set

$$\mathcal{Z} = \{z = [x_1; x_2; w_1; w_2]\},$$

where

$$w_1 = x_1^3, \quad w_2 = x_2^3, \quad \dot{x}_1 = -w_1 - w_2, \quad \dot{x}_2 = x_1 - x_2.$$

The *trivial* IQC for $\mathcal{Z}$ are given by

$$\sigma_{LTI}(z) = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}' P \begin{bmatrix} -w_1 - w_2 \\ x_1 - x_2 \end{bmatrix},$$

where $P = P'$ is an arbitrary symmetric 2-by-2 matrix. Among the non-trivial IQC's valid for $\mathcal{Z}$, the simplest two represent the circle and the Popov criteria, and are defined by

$$\sigma_{NL}(z) = d_1 x_1 w_1 + d_2 x_2 w_2 + q_1 w_1 (-w_1 - w_2) + q_2 w_2 (x_1 - x_2),$$

$$V_\sigma(z(\cdot), t) = 0.25(q_1 x_1(t)^4 + q_2 x_2(t)^4),$$

where $d_k \geq 0$. Let $\tilde{\Sigma}$ be the cone of matrices of the quadratic forms σ . Since we are only proving stability, let $\sigma_0 = 0$. It turns out (and is easy to verify) that the only solutions of the IQC feasibility problem $\sigma \leq 0$ are the ones that make $\sigma = \sigma_{LTI} + \sigma_{NL} = 0$, for example

$$P = \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}, \quad d_1 = d_2 = q_2 = 1, \quad q_1 = 0.$$

The absence of strictly feasible solutions corresponds to the fact that the system is not exponentially stable. Nevertheless, a Lyapunov function candidate can be constructed from the given solution:

$$V(x) = x' P x + 0.25(q_1 x_1^4 + q_2 x_2^4) = 0.5x_1^2 + 0.25x_2^4.$$

This Lyapunov function can be used along the standard lines to prove global asymptotic stability of the equilibrium $x = 0$ in system (8.8),(8.9).

Now consider the case when $\tau \in [0, 0.2]$ is an uncertain parameter. To show that the delayed system (8.8),(8.9) remains stable when $\tau \leq 0.2$, (8.8),(8.9) can be represented by a more elaborate behavior set $\mathcal{Z} = \{z(\cdot)\}$ with

$$z = [x_1; x_2; w_1; w_2; w_3; w_4; w_5; w_6] \in \mathbf{R}^8,$$

satisfying LTI relations

$$\dot{x}_1 = -w_1 - w_2 + w_3, \quad \dot{x}_2 = x_1 - x_2$$

and the nonlinear/infinite dimensional relations

$$w_1(t) = x_1^3, \quad w_2 = x_2^3, \quad w_3 = x_2^3 - (x_2 + w_4)^3,$$

$$w_4(t) = x_2(t - \tau) - x_2(t), \quad w_5 = w_4^3, \quad w_6 = (x_1 - x_2)^3.$$

Some additional IQC are needed to bound the new variables. These will be selected using the perspective of a small gain argument. Note that the perturbation w_4 can easily be bounded in terms of $\dot{x}_2 = x_1 - x_2$. In fact, the LTI system with transfer function $(\exp(-\tau s) - 1)/s$ has a small gain (in almost any sense) when τ is small. Hence a small gain argument would be applicable provided that the gain “from w_4 to $\dot{x}_2$ ” could be bounded as well.

It turns out that the Λ_2 -induced gain from w_4 to $\dot{x}_2$ is unbounded. Instead, we can use the Λ_4 norms. Indeed, the last two components w_5, w_6 of w were introduced in order to handle L_4 norms within the framework of IQC. More specifically, in addition to the trivial IQC with

$$\sigma_{LTI}(z) = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}' P \begin{bmatrix} -w_1 - w_2 + w_3 \\ x_1 - x_2 \end{bmatrix},$$

the set $\mathcal{Z}$ satisfies the IQC $\sigma \geq \dot{V}$, where

$$\begin{aligned} \sigma(z) = & d_1 x_1 w_1 + d_2 x_2 w_2 + q_1 w_1 (-w_1 - w_2 + w_3) + q_2 w_2 (x_1 - x_2) \\ & + d_3 [0.99(x_1 w_1 + x_2 w_2) - x_1 w_3 + 2.5^4 w_4 w_5 - 0.5^4 (x_1 - x_2) w_6] \\ & + q_3 [0.2^4 (x_1 - x_2) w_6 - w_4 w_5], \end{aligned}$$

$d_i \geq 0$. Here the IQC with coefficients d_1, d_2, q_1, q_2 are same as before. The term with d_3 , based on a zero storage function, follows from the inequality

$$0.99(x_1^4 + x_2^4) - x_1(x_2^3 - (x_2 + w_4)^3) + \left(\frac{5w_4}{2}\right)^4 - \left(\frac{x_1 - x_2}{2}\right)^4 \geq 0$$

(which is satisfied for *all* real numbers x_1, x_2, w_4 , and can be checked numerically).

The term with q_3 follows from a gain bound on the transfer function

$$G_\tau(s) = (\exp(-\tau s) - 1)/s$$

from $x_1 - x_2$ to w_4 . It is easy to verify that the Λ_1 norm of its impulse response equals τ , and hence the L^4 induced gain of the causal LTI system with transfer function G_τ will not exceed 1. Consider the function

$$V_d(v(\cdot), T) = -\inf \int_T^\infty \left\{ 0.2^4 |v_1(t)|^4 - \left| \int_{t-\tau}^t v_1(r) dr \right|^4 \right\} dt, \quad (8.10)$$

where the infimum is taken over all functions v_1 which are square integrable on $(0, \infty)$ and such that $v_1(t) = v(t)$ for $t \leq T$. Because of the L_4 gain bound of G_τ with $\tau \in [0, 0.2]$ does not exceed 0.2, the infimum in (8.10) is bounded. Since we can always use $v_1(t) = 0$ for $t > T$, the infimum is non-positive, and hence V_d is non-negative. The IQC defined by the “ q_3 ” term holds with $V_\sigma = q_3 V_d(x_1 - x_2, t)$.

Let

$$\sigma_0(z) = -0.01(x_1 w_1 + x_2 w_2) = -0.01(x_1^4 + x_2^4),$$

which reflects our intention to show that x_1, x_2 will be integrable with fourth power over $(0, \infty)$.

The IQC model cannot be made strictly feasible, but is feasible for

$$P = \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}, \quad d_1 = d_2 = 0.01, \quad d_3 = q_2 = 1, \quad q_1 = 0, \quad q_3 = 2.5^4.$$

A Lyapunov function candidate can be constructed with the help of these P, d_k, q_k :

$$V(x_e(t)) = 0.5x_1(t)^2 + 0.25x_2(t)^4 + 2.5^4 V_d(x_1 - x_2, t),$$

where x_e is the “total state” of the system (in this case, $x_e(T) = [x(T); v_T(\cdot)]$, where $v_T(\cdot) \in L_2(0, \tau)$ denotes the signal $v(t) = x_1(T - \tau + t) - x_2(T - \tau + t)$ restricted to the interval $t \in (0, \tau)$). From the solution of the IQC feasibility problem, it follows that

$$\frac{dV(x_e(t))}{dt} \leq -0.01(x_1(t)^4 + x_2(t)^4).$$

On the other hand, we saw previously that $V(x_e(t)) \geq 0$ is bounded from below. Therefore, $x_1(\cdot), x_2(\cdot) \in \Lambda_4$ (fourth powers of x_1, x_2 are integrable over $(0, \infty)$) as long as the initial conditions are bounded. Thus, the equilibrium $x = 0$ in system (8.8),(8.9) is stable for $0 \leq \tau \leq 0.2$.

9 Model order reduction

9.1 Objectives and challenges of model reduction

Model reduction is one of the most widely encountered “dynamical systems” tasks. In practice, it enables practical use of first principles models for physical phenomena described by partial differential equations, advanced control and signal processing algorithms. Model reduction can also be used to facilitate system identification, data compression, and knowledge extraction.

This class will concentrate on mathematical techniques of model reduction applicable to linear time invariant (LTI) systems, while also venturing into the field of nonlinear models whenever possible. Intentionally simplified, if not trivialized, application examples will be used for motivation and illustration of the theory. The lecture notes presentation will be rigorous, but many formal details will be skipped in the lectures.

9.1.1 Motivating example: the heat equation

Consider the task of modeling the dynamical dependence of time-varying temperature $y = y(t)$ at a point A of a thin non-homogeneous circular wire on the temperature $u = u(t)$ at the opposite point B, where the wire is being forcefully heated/cooled (see Figure 9.27). The relation between $u = u(t)$ and $y = y(t)$ can be written in the form of a

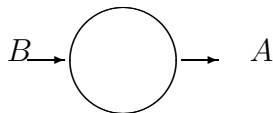


Figure 9.27: Example: the heat equation

partial differential equation

$$\frac{dv(t, \theta)}{dt} = K(\theta) \frac{d^2 v(t, \theta)}{d\theta^2},$$

where $v(t, \theta)$ is temperature of the wire at time t at the point with angular position θ ($\theta = 0$ corresponds to point A), i.e.

$$v(t, -\pi) = v(t, \pi) = u(t), \quad y(t) = v(t, 0),$$

and $K(\theta) > 0$ is a given position-dependent coefficient describing local properties of the wire. This model, while accurate subject to some idealizing assumptions, is not good for

simulation or feedback control design. A model reduction technique would generate a low order LTI system providing an accurate approximation of the true dynamics.

9.1.2 Motivating example: system identification

An important task performed repeatedly by wireless communication devices is “channel identification”, which essentially means finding a good model for electromagnetic signal propagation between two communication points. This can be accomplished by sending a white noise signal through the channel and calculating the statistical spectrum of the signal received. As the next step, this (noisy) spectral data has to be fitted by a low order rational function, as shown on If the order of approximation is allowed to be large

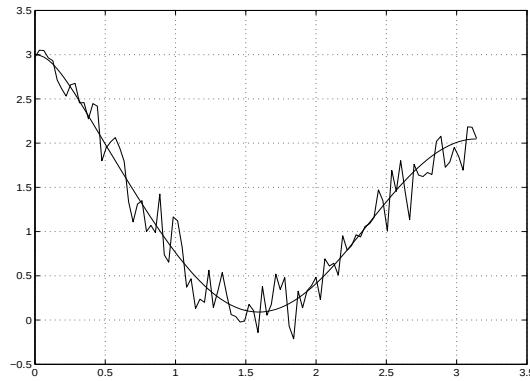


Figure 9.28: Example: noisy data fitting

(to match the number of data points available), the resulting model will try to fit the noise component of the data, which is highly undesirable. An optimal model reduction technique will help to avoid noise fitting by finding the best low order rational approximation.

9.1.3 Simplification of general system models

A general objective of model reduction as a mathematical discipline can be described as that of finding *efficient ways of deriving adequate simplified models of complex systems*.

Here the word *system* refers to a transformation which defines a family of real parameters (*output data*) as a function of another family of real parameters (*input data*). A major object of study in these lectures will be the class of *linear time-invariant dynamical systems*.

A particular technique of model reduction would deal with a specific class of system descriptions by

- specifying a quantitative *system complexity measure* for models from the class;
- defining a notion of accuracy (“adequacy”) for replacing one (“complex”) model with another (“reduced”) one;
- supplying a numerical algorithm which actually performs the reduction;
- presenting evidence (experimental or via a mathematical proof) that the algorithm produces accurate models of low complexity in reasonable time (“algorithm cost”);

Accordingly, to compare two methods of model reduction, one has to take into account accuracy, system complexity reduction, and algorithm cost guarantees associated with them.

9.1.4 Example: reduction of matrix-vector products

Action of a given linear transformation $M : f \mapsto y$ on its unspecified input f is frequently represented as multiplication of a variable real m -vector f by a given real n -by- m matrix M : $y = Mf$. In some applications (image processing, optical simulations), it is desirable to perform such matrix-vector multiplications, where M is known a-priori, as quickly as possible, which typically means minimization of the number of computer operations (multiplication, addition, copying) needed to compute $y = Mf$. It may also be admissible to have a certain amount r of relative error when calculating y quickly.

While the standard definition of a matrix-vector product involves nm operations, some linear transformations $f \mapsto y = Mf$ can be performed much faster, depending on M . Moreover, it could be possible to reduce the minimal number of operations by perturbing M slightly.

Matrix model reduction: a number-of-operations setup The task of finding, for a given M , a faster way of producing an approximation of the product $y = Mf$, can be viewed as a model reduction problem, with the notions of “systems”, “complexity”, “accuracy”, and “efficiency” defined as follows.

Systems: every real n -by- m matrix M defines a *matrix model system* with the set of admissible inputs defined as the set $\mathbf{R}^m = \{f\}$ of all real m -vectors f , and outputs $y \in \mathbf{R}^n$ defined by $y = Mf$.

System complexity: define complexity $\mathcal{N} = \mathcal{N}(M)$ of a matrix model system as the minimal number of binary addition, multiplication, and memory copy operations needed

to implement the input-to-output transformation $f \mapsto y = Mf$. Note that, while it is easy to *define* $\mathcal{N}(M)$ it is usually quite difficult to *calculate* the quantity.

Accuracy: define a numerical measure of error of approximating an original matrix M by a “reduced” matrix $\hat{M}$ as the maximal Euclidean norm $|e|$ of the output matching error vector $e = Mf - \hat{M}f$ when the Euclidean norm of input f is bounded by $|f| \leq 1$. In linear algebra, this quantity is known as the *operator norm* $\|M - \hat{M}\|$, or *largest singular value* $\sigma_{\max}(M - \hat{M})$ of $M - \hat{M}$.

Algorithm cost: define efficiency of a model reduction algorithm as the maximal number of operations needed to produce a reduced model $\hat{M}$ for a given M .

Even after the basic notions of system class, complexity, accuracy, and efficiency are defined, there is a variety of possible approaches to follow. For example, one can formulate an optimal matrix reduction problem (given M and $r > 0$, find $\hat{M} = \hat{M}(M, r)$ such that $\|M - \hat{M}\| \leq r$, and $\mathcal{N}(\hat{M})$ is as small as possible). Alternatively, one can pick a transformation $M \mapsto \mathcal{F}(M)$ and try to prove, either formally or via numerical experiments, that, for most M , $\mathcal{N}(\mathcal{F}(M)) < c\mathcal{N}(M)$, and $\|M - \mathcal{F}(M)\|$ is small. However, despite the number of options available, most of them are likely to end up nowhere, due to the difficulty of working with the “uncomputable” complexity measure $\mathcal{N}(M)$.

Matrix model reduction: a rank setup As it frequently happens, a relatively minor modification of the matrix reduction setup leads to a computationally efficient algorithm.

Let us note that the rank of M defines an upper bound on the minimal number of operations needed to implement transformation $f \mapsto Mf$. Indeed, if $k = \text{rank}(M)$ then $M = VU$ where V is an n -by- k matrix, and U is a k -by- m matrix. Hence, Mf can be found by forming Uf first (km operations) and then calculating VUf (kn operations), which yields a total of $k(n + m)$ operations. When $k \ll n$ and $k \ll m$, this constitutes a substantial reduction in the number of operations.

Thus, it appears to be natural to consider a modified matrix reduction setup, in which $\mathcal{N}(\hat{M})$ is replaced by $\text{rank}(\hat{M})$ as a system complexity measure. This will be referred to as the *matrix rank reduction*. Note that this is not equivalent to the number-of-operations setup, because multiplication by some matrices of full rank can be performed very quickly (for example, multiplication by an upper triangular n -by- n matrix with 1’s above the diagonal can be performed in n steps).

Optimal matrix rank reduction A solution of the optimal matrix rank reduction problem is a standard part of introductory linear algebra. Indeed, let $\sigma_k^2 = \lambda_k$, where $k \in \{1, 2, \dots, m\}$ and $\sigma_k \geq 0$, be the ordered eigenvalues of $M'M$ (i.e. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k$). Then $\|M - \hat{M}\| \geq \sigma_k$ for every matrix $\hat{M}$ of rank less than k . Moreover, a matrix

$\hat{M}$ of rank less than k such that $\|M - \hat{M}\| = \sigma_k$ can be defined by $\hat{M} = VV'MUU'$, where the columns of U are the first $k - 1$ orthonormal eigenvectors of $M'M$ (i.e. those corresponding to eigenvalues $\lambda_1, \dots, \lambda_{k-1}$), and the columns of V are the first $k - 1$ orthonormal eigenvectors of MM' .

Thus, when the accuracy measure is defined as the operator norm, optimal matrix rank reduction can be implemented via the so-called *singular value decomposition*, which is, essentially eigenvalue decomposition of positive semidefinite matrices. It is interesting to note that for many other accuracy measures there is still no efficient solution available for the optimal matrix rank reduction problem.

9.1.5 Challenges of model reduction

There is a number of formulation changes which will complicate the matrix reduction process dramatically. It is instructive to discuss some of these modifications, as they are typical for the general model reduction research area.

Partially defined systems We started with the assumption that all coefficients of M are known precisely. This is frequently not the case: some of the coefficients may be unknown, while a general constraint is placed on their dependence on the index (say, $|M_{i,j} - M_{i,j+1}| \leq 0.01$). Alternatively, a noise factor may be present in *all* coefficient data.

A system which is too large A very large matrix can be defined “analytically”, while being too large to be stored in the memory of a computer. For example, what about a 10^{20} - 10^{20} matrix M with entries

$$M_{ij} = \frac{1}{2 + i^2 + j^2 + \cos(i)}?$$

Non-linear and uncertain systems Already a very simple type of nonlinearity – parameter dependence – makes a model reduction problem much harder. Model reduction of more general nonlinear systems remains largely an uncharted territory.

How to compare model reduction methods? Model reduction methods have to be compared with respect to many parameters (accuracy, algorithm cost, type of complexity measure, etc.) An algorithm which is optimal accuracy-wise may be prohibitively expensive, and a cheap algorithm may produce extremely inadequate reduced models. Moreover, theoretical proofs of performance are rare.

9.2 Model reduction by projection

A projection method can be viewed as application of a “lossy” compression to a system’s state, and re-writing equations for the state’s dynamics in terms of a compressed representation. Depending on a system and model class (linear vs. non-linear, state-space vs. input/output, time-varying vs. time-invariant, etc.), different implementations of this approach become typical.

9.2.1 Projection of finite order state space LTI models

The sequence of operations for obtaining a reduced CT LTI SS model

$$\frac{d}{dt}\hat{x}(t) = \hat{A}\hat{x}(t) + \hat{B}f(t), \quad \hat{y}(t) = \hat{C}\hat{x}(t) + \hat{D}f(t)$$

from the original higher-order model

$$\frac{d}{dt}x(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t)$$

can be described in the following way.

Step 1: apply an invertible coordinate change $x(t) = S\bar{x}(t)$, where S is an invertible square matrix, to re-write the original system equations in the form

$$\frac{d}{dt}\bar{x}(t) = \bar{A}\bar{x}(t) + \bar{B}f(t), \quad y(t) = \bar{C}\bar{x}(t) + \bar{D}f(t),$$

where

$$\bar{A} = S^{-1}AS, \quad \bar{B} = S^{-1}B, \quad \bar{C} = CS, \quad \bar{D} = D.$$

Step 2: partition the new state vector $\bar{x}(t)$ as

$$\bar{x}(t) = \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \end{bmatrix},$$

where the dimension of $\bar{x}_1(t)$ equals the desired order of the reduced system; partition the $\bar{A}, \bar{B}, \bar{C}$ matrices accordingly:

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix}, \quad \bar{C} = \begin{bmatrix} \bar{C}_1 & \bar{C}_2 \end{bmatrix}.$$

Step 3: define the reduced system by

$$\hat{A} = \bar{A}_{11}, \quad \hat{B} = \bar{B}_1, \quad \hat{C} = \bar{C}_1, \quad \hat{D} = D.$$

The following reasoning stays usually behind such series of manipulations. At *step 1*, the system states are re-arranged to place the “most important” ones as the first few components of the new state vector $\bar{x}(t)$. At *step 2*, one decides which components of $\bar{x}(t)$ are to be “ignored” (set to zero) in the new equations. *Step 3* defines the resulting simplified state equations.

Selecting a “good” initial transformation S , as well as determining how many states of $\bar{x}(t)$ to keep, is what defines a particular projection-based model reduction algorithm. One can invest a lot of effort in finding a good S , and come up with a high quality reduced model. Alternatively, one can obtain S according to some “cheap” strategy, to get a barely adequate reduced model. Actually, it can be shown that, by selecting an arbitrary S , one can obtain an *arbitrary* (subject to some non-essential restrictions) reduced order system from a given model.

For example, an unobservable CT LTI SS model (i.e. the one with the observability matrix M_o not having full rank) can be transformed to a form with

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & 0 \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix}, \quad \bar{C} = [\bar{C}_1 \quad 0].$$

In this case the reason to consider the state $\bar{x}_2(t)$ “unimportant” is its unobservability, and the projection reduced model

$$\hat{A} = \bar{A}_{11}, \quad \hat{B} = \bar{B}_1, \quad \hat{C} = \bar{C}_1, \quad \hat{D} = D$$

has a transfer matrix which is identical to that of the original system.

As another example, consider the stable causal LTI system defined by

$$\begin{aligned} \dot{x}_1 &= x_1 + x_2 + f, \\ \dot{x}_2 &= -3x_1 - 2x_2, \\ y &= x_1. \end{aligned}$$

When the initial state transformation $x = S\bar{x}$ is defined by the identity matrix $S = I = I_2$, the reduced first order system has transfer function $1/(s-1)$ – very little in common with the original system.

9.2.2 An alternative interpretation of projection MOR

For efficient numerical calculations, as well as for the sake of formal mathematical manipulations, a different (though equivalent) interpretation of the projection approach can be introduced.

Let n be the order of the original model, and let $r < n$ be the desired reduced system order. A particular projection can be defined by specifying an n -by- r matrix V and an r -by- n matrix U such that

$$UV = I_r.$$

The reduced system is then defined by substituting $x(t) = V\hat{x}(t)$ into the original equations, and by post-multiplying the resulting differential equation by U on the left, which yields

$$\hat{A} = UAV, \quad \hat{B} = UB, \quad \hat{C} = CV, \quad \hat{D} = D.$$

There are two important messages about this representation of a projection MOR framework. First, the outcome of the procedure cannot be changed much by replacing columns of V with their linear combinations, as well as by replacing rows of U by their linear combinations, as follows from the next statement.

Theorem 9.1 *If matrices V_1, V_2, U_1, U_2 are such that*

- (a) $U_1V_1 = U_2V_2 = I_r$,
- (b) *the linear spans of the columns of V_1 and V_2 are identical, and*
- (c) *the linear spans of the rows of U_1 and U_2 are identical,*

then the systems

$$\hat{G}_1 := \left(\frac{U_1AV_1}{CV_1} \middle| \frac{U_1B}{D} \right), \quad \hat{G}_2 := \left(\frac{U_2AV_2}{CV_2} \middle| \frac{U_2B}{D} \right)$$

have identical transfer matrices.

Proof. According to (b), there exists an r -by- r matrix S_v such that $V_2 = V_1S_v$. Similarly, according to (c), there exists an r -by- r matrix S_u such that $U_2 = S_uU_1$. Hence, from (a), $S_uS_v = I$, i.e. $S_u = S_v^{-1}$, and the state space model of $\hat{G}_2$ is obtained from that of $\hat{G}_1$ by replacing its state $\hat{x}_1$ with $\hat{x}_2 = S_u\hat{x}_1$. Since invertible linear transformations of the state vector do not change transfer matrices, the proof is complete. ■

In particular, without loss of generality, one can limit attention only to those matrices V for which $V'V = I_r$. Similarly, one can limit attention only to those matrices U for which $UU' = I_r$, though in this case it would be restrictive to assume that $V'V = I_r$ as well.

To relate the two approaches to projection MOR, note that a pair of matrices U, V satisfying the condition $UV = I_r$ can always be complemented to a pair of mutually inverse matrices

$$S_u = \begin{bmatrix} U \\ \Delta_u \end{bmatrix}, \quad S_v = \begin{bmatrix} V & \Delta_v \end{bmatrix}.$$

If

$$S_v^{-1}AS_v = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix}, \quad S_v^{-1}B = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix}, \quad CS_v = \begin{bmatrix} \bar{C}_1 & \bar{C}_2 \end{bmatrix},$$

where $\bar{A}_{11}, \bar{B}_1, \bar{C}_1$ have appropriate dimensions, we have

$$\bar{A}_{11} = UAV, \quad \bar{B}_1 = UB, \quad \bar{C}_1 = CV,$$

i.e. the two projections yield same outcome.

Similarly one can relate the original framework to the coordinate-free one by defining

$$V = S \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \quad U = \begin{bmatrix} I_r & 0 \end{bmatrix} S^{-1},$$

where $x = S\bar{x}$ is the original coordinate transformation.

9.2.3 Projections for other model types

For *discrete time LTI SS models*, projection MOR is defined in exactly the same way as in the continuous time case:

$$x[k+1] = Ax[k] + Bf[k], \quad y[k] = Cx[k] + Df[k]$$

is reduced to

$$\hat{x}[k+1] = UAV\hat{x}[k] + UBf[k], \quad y[k] = CV\hat{x}[k] + Df[k],$$

where $UV = I$.

For *descriptor system models*, the modification is straightforward:

$$E\dot{x}(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t)$$

is replaced with

$$UEV \frac{d}{dt} \hat{x}(t) = UAV \hat{x}(t) + UBf(t), \quad y(t) = CV \hat{x}(t) + Df(t),$$

where the constraint $UV = I$ is dropped, but instead a requirement that $\det(sUEV - UAV)$ is not identically zero is imposed, provided that $\det(sE - A)$ is not identically zero.

For *nonlinear time-invariant state space models*

$$\dot{x}(t) = a(x(t), f(t)), \quad y(t) = h(x(t), f(t)),$$

the assumed transformation from $\hat{x}(t)$ to $x(t)$ and back should be allowed to become nonlinear, i.e.

$$x(t) = V(\hat{x}(t)), \quad \hat{x}(t) = U(x(t)),$$

where $V : \mathbf{R}^r \mapsto \mathbf{R}^n$ and $U : \mathbf{R}^n \mapsto \mathbf{R}^r$ are given differentiable functions satisfying the identity

$$U(V(z)) = z \quad \forall z \in \mathbf{R}^r.$$

The resulting model is

$$\frac{d}{dt} \hat{x}(t) = \dot{U}(V(\hat{x}(t)))a(V(\hat{x}(t)), f(t)), \quad y(t) = h(V(\hat{x}(t)), f(t)).$$

While the number of state gets reduced this way, it is not obvious (and is not proven to any degree of generality) that the number of arithmetic operations needed to evaluate $d\hat{x}(t)/dt$ is smaller than the number of operations needed to evaluate $dx(t)/dt$.

For *linear time-varying models*

$$\dot{x}(t) = A(t)x(t) + B(t)f(t), \quad y(t) = C(t)x(t) + D(t)f(t),$$

the projected reduced model is frequently defined as

$$\frac{d}{dt} \hat{x}(t) = U(t)(A(t)V(t) - \dot{V}(t))\hat{x}(t) + U(t)B(t)f(t), \quad y(t) = C(t)V(t)\hat{x}(t) + D(t)f(t),$$

where U, V are differentiable functions of time such that $U(t)V(t) = I$ at all moments. Once again, while reduction in the dimension of the state space is indisputable, the situation with complexity of reduced model simulation is not as clear, since the new coefficients $\hat{A}(t), \hat{B}(t), \hat{C}(t)$ may be less regular functions of time, which usually requires extra effort in simulation.

For *general causal input/output models*, where an explicit state description may be unavailable, the *past input/output history*

$$x(t) = (f(\tau)|_{\tau \leq t}, y(\tau)|_{\tau \leq t})$$

can serve as a replacement. Then the reduced state can be defined by a transformation

$$\hat{x}(t) = U(x(t)) = U(f(\tau)|_{\tau \leq t}, y(\tau)|_{\tau \leq t}).$$

9.2.4 Preservation of transfer matrix moments

Consider an n -th order state space LTI model

$$\frac{d}{dt}x(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t), \quad (9.1)$$

with transfer matrix

$$G(s) = D + C(sI_n - A)^{-1}B,$$

and its projection reduced r -th model

$$\frac{d}{dt}\hat{x}(t) = UAV\hat{x}(t) + UBf(t), \quad \hat{y}(t) = CV\hat{x}(t) + Df(t), \quad (9.2)$$

with transfer matrix

$$\hat{G}(s) = D + CV(sI_r - UAV)^{-1}UB,$$

where U, V are matrices of dimensions r -by- n and n -by- r respectively, such that $UV = I_r$, and $r < n$. As it was mentioned before, no system property of importance (such as stability, passivity, transfer matrix values, etc.) is guaranteed to be preserved by a generic projection of this type. However, when the projection matrices U, V are properly defined, some features of the original system can be transferred to the reduced one.

This subsection provides *sufficient* conditions for preservation of *transfer matrix moments* under projection model reduction. Here by the *moments* of a transfer matrix $H = H(s)$ at a given point s_0 , which is *not* a pole of $H(\cdot)$, we mean the values $H^{(k)}(s_0)$ of its derivatives and its own value $H(s_0) = H^{(0)}(s_0)$ at $s = s_0$.

Theorem 9.2 *Consider system (6.17) and its projection reduced model (6.18). Let $s_0 \in \mathbf{C}$ be a complex number which is not an eigenvalue of A or UAV .*

- (a) *If, for some column vector $\bar{f}$, $(s_0I_n - A)^{-k-1}B\bar{f}$ belongs to the set $\mathcal{R}(V)$ of all linear combinations of columns of V for all $k = 0, \dots, N$ then*

$$G^{(k)}(s_0)\bar{f} = \hat{G}^{(k)}(s_0)\bar{f} \quad \text{for } k = 0, \dots, N.$$

(b) If $\bar{q}C(s_0I_n - A)^{-k-1}$ is a linear combination of the rows of U for some row vector $\bar{q}$ for all $k = 0, \dots, N$ then

$$\bar{q}G^{(k)}(s_0) = \bar{q}\hat{G}^{(k)}(s_0) \quad \text{for } k = 0, \dots, N.$$

This observation is frequently used in determining the projection matrices U, V : the rows of U and the columns of V are chosen as bases of the linear subspaces spanned by the real and imaginary parts of the rows of

$$C(j\omega_i^c I_n - A)^{-k-1}, \quad k = 0, \dots, N_i^c,$$

and the columns of

$$(j\omega_i^b I_n - A)^{-k-1}B, \quad k = 0, \dots, N_i^b,$$

respectively, where ω_i^c and ω_i^b are selected “important” frequencies, and the maximal powers N_i^c, N_i^b indicate the degree of accuracy desired at ω_i^c, ω_i^b . The resulting projection algorithms provide a numerically robust implementation of the *moments matching* approach to model reduction, to be discussed in future lectures.

Proof. Without loss of generality, assume that $D = 0$. To prove (a), note that

$$G^{(k)}(s_0) = k!C(s_0I_n - A)^{-k-1}B \quad (k = 0, 1, \dots),$$

and

$$\hat{G}^{(k)}(s_0) = k!CV(s_0I_r - UAV)^{-k-1}UB \quad (k = 0, 1, \dots).$$

Note that the vectors

$$x_k = (s_0I_n - A)^{-k-1}B\bar{f}, \quad (k = 0, 1, \dots)$$

are uniquely defined by the recursive linear equations

$$s_0x_0 = Ax_0 + B\bar{f}, \quad s_0x_k = Ax_k + x_{k-1}, \quad (k = 1, \dots).$$

By the assumption, $x_k = V\hat{x}_k$ for $k = 0, \dots, N$ for some vectors $\hat{x}_k$. Hence

$$s_0V\hat{x}_0 = AV\hat{x}_0 + B\bar{f}, \quad s_0V\hat{x}_k = AV\hat{x}_k + V\hat{x}_{k-1}, \quad (k = 1, \dots, N).$$

Multiplying these equalities by U on the left yields

$$s_0\hat{x}_0 = \hat{A}\hat{x}_0 + \hat{B}\bar{f}, \quad s_0\hat{x}_k = \hat{A}\hat{x}_k + \hat{x}_{k-1}, \quad (k = 1, \dots, N),$$

where $\hat{A} = UAV$ and $\hat{B} = UB$. Hence

$$\hat{x}_k = (s_0I_r - \hat{A})^{-k-1}\hat{B}\bar{f}, \quad (k = 0, \dots, N),$$

which in turn yields

$$C(s_0 I_n - A)^{-k-1} B \bar{f} = Cx_k = CU\hat{x}_k = \hat{C}\hat{x}_k = \hat{C}(s_0 I_r - \hat{A})^{-k-1} \hat{B} \bar{f}$$

for $k = 0, \dots, N$, thus proving (a).

The proof of (b) is similar, and uses

$$p_k = \bar{q}C(s_0 I_n - A)^{-k-1}$$

in place of x_k . By the assumption, $p_k = \hat{p}_k U$ for $k = 0, \dots, N$ for some row vectors $\hat{p}_k$. Hence

$$s_0 \hat{p}_0 U = \hat{p}_0 U A + \bar{q}C, \quad s_0 \hat{p}_k U = \hat{p}_k U A + \hat{p}_{k-1} U, \quad (k = 1, \dots, N).$$

Multiplying these equalities by V on the right yields

$$s_0 \hat{p}_0 = \hat{p}_0 \hat{A} + \bar{q} \hat{C}, \quad s_0 \hat{p}_k = \hat{p}_k \hat{A} + \hat{p}_{k-1}, \quad (k = 1, \dots, N),$$

where $\hat{C} = CV$. Hence

$$\hat{p}_k = \bar{q}C(s_0 I_r - \hat{A})^{-1}, \quad (k = 0, \dots, N),$$

which yields

$$\bar{q}C(s_0 I_n - A)^{-k-1} B = p_k B = \hat{p}_k U B = \hat{p}_k \hat{B} = \bar{q} \hat{C}(s_0 I_r - \hat{A})^{-k-1} \hat{B}$$

for $k = 0, \dots, N$, thus proving (b). ■

The results of this subsection apply equally to discrete time LTI state space models. They can also be extended to the case of descriptor models, to yield the following statement, which is proven by a similar argument.

Theorem 9.3 *Consider the descriptor system model*

$$E \frac{d}{dt} x(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t),$$

and its projection reduced model

$$\hat{E} \frac{d}{dt} \hat{x}(t) = \hat{A} \hat{x}(t) + \hat{B} f(t), \quad \hat{y}(t) = \hat{C} \hat{x}(t) + Df(t),$$

where

$$\hat{A} = UAV, \quad \hat{E} = UEV, \quad \hat{B} = UB, \quad \hat{C} = CV,$$

and U, V are matrices of dimensions r -by- n and n -by- r respectively, such that $r < n$. Let $s_0 \in \mathbf{C}$ be a complex number such that both matrices $s_0 E - A$ and $s_0 \hat{E} - \hat{A}$ are invertible

(a) If, for some column vector $\bar{f}$, $[(s_0E - A)^{-1}E]^k(s_0E - A)^{-1}B\bar{f}$ belongs to the set $\mathcal{R}(V)$ of all linear combinations of columns of V for all $k = 0, \dots, N$ then

$$G^{(k)}(s_0)\bar{f} = \hat{G}^{(k)}(s_0)\bar{f} \quad \text{for } k = 0, \dots, N,$$

where

$$G(s) = C(sE - A)^{-1}B + D, \quad \hat{G}(s) = \hat{C}(s\hat{E} - \hat{A})^{-1}\hat{B} + \hat{D}.$$

(b) If $\bar{q}C(s_0E - A)^{-1}[E(s_0E - A)^{-1}]^k$ is a linear combination of the rows of U for some row vector $\bar{q}$ for all $k = 0, \dots, N$ then

$$\bar{q}G^{(k)}(s_0) = \bar{q}\hat{G}^{(k)}(s_0) \quad \text{for } k = 0, \dots, N.$$

9.2.5 Stability preservation

In most applications, a reduced model of a stable system is required to be stable. However, stability preservation does not come automatically in projection model reduction.

The following result provides a sufficient condition for stability of projection reduced systems in the continuous time case. Remember that a controllable and observable state space model (6.17) defines a stable causal system if and only if A is a Hurwitz matrix (all eigenvalues have negative real part), or, equivalently, if there exists a symmetric strictly positive definite matrix $P = P' > 0$ such that the *Lyapunov equality*

$$PA + A'P = -R \tag{9.3}$$

is satisfied for some positive semidefinite symmetric matrix $R \geq 0$ such that the pair (A, R) is controllable (note that the controllability is implied, in particular, when R is strictly positive definite). Such $P = P'$ defines an energy-like quantity $V(x(t)) = x(t)'Px(t)$ which is guaranteed not to increase when the external signal $f = f(t)$ equals zero. In many applications, $P = P'$ is readily available from the physical laws of energy conservation and dissipation.

Theorem 9.4 *Consider system (6.17) and its projection reduced model (6.18). Let $P = P'$ and V be such that*

$$V'PV > 0, \quad V'(PA + A'P)V < 0.$$

If U is defined by

$$U = (V'PV)^{-1}V'P$$

then $\hat{A} = UAV$ is a Hurwitz matrix satisfying the Lyapunov inequality

$$\hat{P}\hat{A} + \hat{A}'\hat{P} < 0 \quad \text{for } \hat{P} = V'PV.$$

Proof. The proof is based on a straightforward verification of the identity

$$\hat{P}\hat{A} = V'PAV.$$

■

The theorem suggests that stability of the reduced system can be guaranteed easily by limiting the freedom of selecting U for a given V : to use the result, one has U uniquely defined by V and a conserved energy measure matrix P . Note that condition $V'(PA + A'P)V < 0$ can be replaced by the weaker one:

$$V'(PA + A'P)V = -\hat{R},$$

where $\hat{R} = \hat{R}' \geq 0$ and the pair $(\hat{A}, \hat{R})$ is controllable.

For descriptor models, similar results are available. Remember that a certificate of (marginal) stability for a descriptor system

$$E \frac{d}{dt} x(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t) \quad (9.4)$$

is given by a symmetric matrix $P = P'$ such that $E'PE \geq E'E$ and

$$E'PA + A'PE \leq 0. \quad (9.5)$$

Theorem 9.5 *Let $P = P'$ be a symmetric n -by- n matrix and V be an n -by- r matrix such that*

$$\hat{P} = V'E'PEV > 0 \quad \text{and} \quad V'(E'AP + PA'E')V < 0.$$

Let

$$U = (V'E'PEV)^{-1}V'E'P.$$

Then the reduced system

$$\hat{E} \frac{d}{dt} \hat{x}(t) = \hat{A}x(t) + \hat{B}f(t), \quad \hat{y}(t) = \hat{C}\hat{x}(t) + Df(t),$$

where

$$\hat{E} = UEV = I_r, \quad \hat{A} = UAV, \quad \hat{B} = UB, \quad \hat{C} = CV,$$

is stable, and has a Lyapunov function

$$\hat{V}(\hat{x}(t)) = \hat{x}(t)' \hat{P} \hat{x}(t)$$

which decreases monotonically for system solutions with $f(t) = 0$.

Proof. The proof of this theorem is a straightforward verification of the identity

$$\hat{P}\hat{A} = V'E'PAV.$$

■

Similar stability preservation results are also available for discrete time models

$$x[k+1] = Ax[k] + Bf[k], \quad y[k] = Cx[k] + Df[k], \quad (9.6)$$

where the stability of a minimal state space model (9.6) is equivalent to A being a Schur matrix (all eigenvalues strictly inside the unit disc), and a stability certificate is usually given in the form of a symmetric positive definite matrix $P = P' > 0$ such that the *discrete time Lyapunov inequality*

$$P - A'PA > 0$$

is satisfied.

Theorem 9.6 *Consider a discrete time LTI model (9.6). Let $P = P' \geq 0$ and V be matrices of dimensions n -by- n and n -by- r respectively, such that*

$$V'PV > 0, \quad V'(A'PA - P)V < 0.$$

If U is defined by

$$U = (V'PV)^{-1}V'P$$

then $\hat{A} = UAV$ is a Schur matrix satisfying the Lyapunov inequality

$$\hat{A}'\hat{P}\hat{A} - \hat{P} < 0 \quad \text{for} \quad \hat{P} = V'PV,$$

and hence the projection reduced model

$$\hat{x}[k+1] = UAV\hat{x}[k] + UBf[k], \quad \hat{y}[k] = CV\hat{x}[k] + Df[k]$$

is stable.

Proof. The proof is based on a straightforward verification of the inequality

$$PV(V'PV)^{-1}V'P \leq P.$$

In fact, the assumption $V'(A'PA - P)V < 0$ in the theorem formulation can be relaxed to the weaker one

$$V'PV - V'A'PV(V'PV)^{-1}V'PAV > 0.$$

■

9.2.6 L2 gain and passivity preservation

Let us call a stable LTI system model with a proper rational transfer matrix $H = H(s)$ *contracting* if

$$\sigma_{\max}(H(j\omega)) = \|H(j\omega)\| < 1 \quad \forall \omega \in \mathbf{R}.$$

A formally different, but actually very similar definition concerns *passivity*. A stable LTI system model with a proper *square* rational transfer matrix $H = H(s)$ is called *passive* if

$$H(j\omega) + H(j\omega)' < 0 \quad \forall \omega \in \mathbf{R},$$

where the prime sign denotes Hermitian conjugation.

Both contractivity and passivity serve as frequency domain conditions for *input/output energy conservation*, having an interpretation that one gets less energy *out* of a system than the amount that was put *into* the system. In circuits applications, for example, contractivity applies to voltage in/voltage out subsystems with large input resistance and small output resistance, so that the supplied/extracted energies equal the integrals of input/output squared, while passivity applies to single-port blocks with voltage and current serving as input and output, so that the difference between supplied and extracted energies equals the integral of the input/output product. Contractivity and passivity are important because feedback interconnections of energy preserving systems preserve the total energy, which implies stability.

As it can be expected, all cases of passivity and contractivity of state space models come with a positive definite quadratic Lyapunov function representing a conserved (dissipated) energy quantity. Mathematically, such statements are special cases of the *Kalman-Yakubovich-Popov Lemma*, which will be presented here without a proof and with a simplifying assumption of strict properness of $H(s)$.

Theorem 9.7 *Let*

$$\dot{x}(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t)$$

be a minimal model of an n -th order stable strictly proper causal system with transfer matrix $H = H(s)$. Then

- (a) *$H = H(s)$ is contracting if and only if there exists a symmetric positive definite matrix $P = P' > 0$ such that the Riccati inequality*

$$PA + A'P + C'C + PBB'P < 0$$

is satisfied;

(b) $H = H(s)$ is passive if and only if there exists a symmetric positive definite matrix $P = P' > 0$ such that

$$C = B'P \quad \text{and} \quad PA + A'P < 0.$$

The matrix inequality from (a) means that

$$\frac{dV(x(t))}{dt} \leq |f(t)|^2 - |y(t)|^2$$

for all solutions of the system equations, where

$$V(\bar{x}) = \bar{x}'P\bar{x} \geq 0$$

is the “potential energy” of the system. Similarly, the matrix inequality from (b) means that

$$\frac{dV(x(t))}{dt} \leq 2f(t)'y(t).$$

It turns out that the same way of selecting U when P and V are given, that was used to preserve stability in projection model reduction, can be employed to preserve contractivity and passivity as well.

Theorem 9.8

(a) If $P = P'$ is such that

$$\hat{P} = V'PV > 0 \quad \text{and} \quad V'(PA + A'P + C'C + PBB'P)V < 0,$$

then the projection reduced model

$$\frac{d}{dt}\hat{x}(t) = \hat{A}\hat{x}(t) + \hat{B}f(t), \quad \hat{y}(t) = \hat{C}\hat{x}(t), \quad (9.7)$$

where

$$\hat{A} = UAV, \quad \hat{B} = UB, \quad \hat{C} = CV, \quad U = (V'PV)^{-1}V'P,$$

is contracting, and its coefficients satisfy the Riccati inequality

$$\hat{P}\hat{A} + \hat{A}'\hat{P} + \hat{C}'\hat{C} + \hat{P}\hat{B}\hat{B}'\hat{P} < 0.$$

(b) If $P = P'$ is such that

$$CV = B'PV, \quad \hat{P} = V'PV > 0 \quad \text{and} \quad V'(PA + A'P)V < 0$$

then the projection reduced model (9.7) is passive, and its coefficients satisfy

$$\hat{C} = \hat{B}'\hat{P} \quad \text{and} \quad \hat{P}\hat{A} + \hat{A}'\hat{P} < 0.$$

9.3 Balanced Truncation

Balanced truncation is an important projection model reduction method which delivers high quality reduced models by making an extra effort in choosing the projection subspaces.

9.3.1 Motivation: removal of (almost) uncontrollable/unobservable modes

Removing an unobservable or an uncontrollable mode is an easy way of reducing the dimension of the state vector in a state space model. For example, system

$$\begin{aligned}\dot{x}_1 &= -x_1 + x_2 + f, \\ \dot{x}_2 &= -2x_2, \\ y &= x_1 + x_2,\end{aligned}$$

can be replaced by

$$\dot{x} = -x + f, \quad y = x$$

without changing its transfer function. In this state space model, with

$$A = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad D = 0,$$

the controllability matrix

$$M_c = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$$

satisfies $pM_c = 0$ for $p = \begin{bmatrix} 0 & 1 \end{bmatrix}$, and hence the variable $px = x_2$ represents an *uncontrollable* mode. The removal of such mode can be viewed as a canonical projection model reduction

$$A \mapsto \hat{A} = UAV, \quad B \mapsto \hat{B} = UB, \quad C \mapsto \hat{C} = CV,$$

where the columns of V form a basis in the column range of M_c (which is the same as the null space of p), and U can be selected quite arbitrarily subject to the usual constraint $UV = I$.

Strictly speaking, the example above cannot even be considered as “model reduction”, as the orders of the original and the projected systems are both equal to 1. A more interesting situation is represented by the perturbed model

$$\begin{aligned}\dot{x}_1 &= -x_1 + x_2 + f, \\ \dot{x}_2 &= -2x_2 + \epsilon f, \\ y &= x_1 + x_2,\end{aligned}$$

(same A, C, D but a modified B), where $\epsilon > 0$ is a parameter. Intuitively, one can expect that, when $\epsilon > 0$ is small enough,

$$\dot{x} = -x + f, \quad y = x$$

is still a good reduced model. This expectation can be related to the fact that x_2 is difficult to control by f when $\epsilon > 0$ is small. One can say that the *mode* $x_2 = px$, which corresponds to the left (row) eigenvector of the A matrix ($pA = -2p$ in this case), is almost uncontrollable, which can be seen directly from the transfer function

$$G(s) = \frac{s + 2 + \epsilon}{(s + 2)(s + 1)} = \frac{1 + \epsilon}{s + 1} - \frac{\epsilon}{s + 2},$$

which has a small coefficient at $(s + 2)^{-1}$ in its partial fraction expansion.

One can attempt to introduce a measure of importance of an LTI system's mode (understood as a pole a of system's transfer function $G(s)$) as something related to the absolute value of the coefficient with which $(s - a)^{-1}$ enters the partial fraction expansion of G . However, it is rarely a good idea to base a model reduction algorithm solely on removal of “unimportant” system modes. For example, both modes $a = 1 - \epsilon$ and $a = 1 + \epsilon$ of the LTI system with transfer function

$$H(s) = \frac{1}{s + 1 - \epsilon} + \frac{1}{s + 1 + \epsilon},$$

where $\epsilon > 0$ is small, are equally important, and none can be removed without introducing a huge model reduction error, despite the fact that a very good reduced model $\hat{H}$ is given by

$$\hat{H}(s) = \frac{2}{s + 1}.$$

Balanced truncation is based on introducing a special joint measure of controllability and observability for *every* vector in the state space of an LTI system. Then, the reduced model is obtained by removing those components of the state vector which have the lowest importance factor in terms of this measure.

9.3.2 Observability measures

Consider a state space model

$$\dot{x}(t) = Ax(t) + Bf(t), \quad y(t) = Cx(t) + Df(t), \quad (9.8)$$

where A is an n -by- n Hurwitz matrix (all eigenvalues have negative real part). When $f(t) \equiv 0$ for $t \geq 0$, the value of the output $y(t)$ at a given moment t is uniquely defined by $x(0)$, and converges to zero exponentially as $t \rightarrow +\infty$. Hence the integral

$$E_o = \int_0^\infty |y(t)|^2 dt,$$

measuring the “observable output energy” accumulated in the initial state, is a function of $x(0)$, i.e. $E_o = E_o(x(0))$. Moreover, since $y(t)$ is a *linear* function of $x(0)$, E_o will be a *quadratic form* with respect of $x(0)$, i.e.

$$E_o(x(0)) = x(0)' W_o x(0)$$

for some symmetric matrix W_o .

The quadratic form $E_o = E_o(x(0))$ can be used as a *measure of observability* defined on the state space of system (9.8). In particular, $E_o(x(0)) = 0$ whenever

$$M_o x(0) = [C; CA; \dots; CA^{n-1}] x(0) = 0.$$

The positive semidefinite matrix W_o of the quadratic form E_o is called the *observability Gramian*³. Since, by the definition, $E_o(x(0)) \geq 0$ for all $x(0)$, the matrix W_o is always positive semidefinite. Indeed, $W_o > 0$ whenever the pair (C, A) is observable.

It is important to notice that the word “system state” actually includes two different meanings. One meaning is that of a *primal* state, which, for a general model (9.8), is a column n -vector. For example, for a state space model

$$\dot{x}_1 = -x_1 + f, \quad \dot{x}_2 = -x_2 + f, \quad y = x_1 + x_2, \quad (9.9)$$

a primal state is a particular column vector value of $x(t) \in \mathbf{R}^2$, such as $x(-1.2) = [1; -7]$. Such column vector values are more precisely referred to as the *primal* states of (9.9).

On the other hand, one frequently refers to (9.9) as a “system with two states” (despite the fact that the set $\mathbf{R}^2$ has an infinite number of elements), and, in this context, $x_1 = x_1(t)$ and $x_2 = x_2(t)$ can be referred to as the two states of the system. Let us call such states the *dual* states. For a general model (9.8), a dual state is a particular linear combination $x_p = p x(t)$ of the scalar components of $x = x(t)$, defined by a row n -vector p . For example, in (9.9), the dual states $x_1 = x_1(t)$ and $x_2 = x_2(t)$ are defined by row vectors $p = [1 \ 0]$ and $p = [0 \ 1]$ respectively.

³Whether this should be spelled as “Grammian” or “Gramian”, is unclear

Therefore, it is natural to ask for a definition of an observability measure of a given *dual* state of (9.8). It is defined as

$$E^o(p) = \inf_{x_0: px_0=1} E_o(x_0) \text{ for } p \neq 0,$$

i.e. as the minimal output energy which can be observed for $t \geq 0$ when $px(0) = 1$. Note that infimum over an empty set equals plus infinity, hence $E^o(0) = \infty$. When the pair (C, A) is observable, and hence $W_o > 0$ is invertible, the dual observability measure is given by

$$E^o(p) = \frac{1}{pW_o^{-1}p'} \text{ for } p \neq 0.$$

The following theorem is frequently utilized for computing W_o numerically.

Theorem 9.9 *W_o is the unique solution of the Lyapunov equation*

$$W_o A + A' W_o + C' C = 0. \tag{9.10}$$

Proof. Since system (9.8) is time invariant, the identity

$$x(0)' W_o x(0) = \int_0^\infty |Cx(\tau)|^2 d\tau$$

implies

$$x(t)' W_o x(t) = \int_t^\infty |Cx(\tau)|^2 d\tau.$$

Differentiating the second identity with respect to t at $t = 0$ yields

$$2x(0)' W_o A x(0) = -|Cx(0)|^2$$

for all $x(0) \in \mathbf{R}^n$. Comparing the coefficients on both sides of the quadratic identity yields (9.10). ■

Finding a numerical solution of (9.10) is not easy when n is about 10^4 and larger. In such situation, Theorem 9.9 can be used as a basis for finding an approximation of W_o .

It is important to understand that the observability measure alone should not be the only numerical test for choosing which states to eliminate in a model reduction procedure. Instead, a combination of observability *and* a controllability measures, to be introduced in the next subsection, should be used.

9.3.3 Controllability measures

Since A is a Hurwitz matrix, every input signal $f = f(t)$ of finite energy, i.e. such that

$$\|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty,$$

corresponds to a unique initial condition $x(0)$ in (9.8) for which the corresponding solution $x = x(t)$ satisfies $x(t) \rightarrow 0$ as $t \rightarrow -\infty$. This solution is given by

$$x(t) = \int_0^{\infty} e^{A\tau} B f(t - \tau) d\tau,$$

where e^M denotes the *matrix exponent* of a square matrix M . One can say that input $f = f(t)$ *drives* the system state from $x(-\infty) = 0$ to $x(0) = X(f(\cdot))$.

Let p be a 1-by- n row vector, so that the product $px(t)$ is a dual state of (9.8) – a linear combination of components of the state space vector. The (dual) controllability measure $E^c = E^c(p)$ is defined as the maximal value of $|px(0)|^2$ which can be achieved by using an input $f = f(t)$ of unit energy:

$$E^c(p) = \max\{|pX(f(\cdot))|^2 : \|f\| \leq 1\}.$$

Accordingly, the *primal* controllability measure $E_c = E_c(x_0)$ is defined for a column vector $x_0 \in \mathbf{R}^n$ as

$$E_c(x_0) = \inf_{p: px_0=1} E^c(p).$$

The following statement describes some basic properties of these controllability measures.

Theorem 9.10 *Assuming that A is an n -by- n Hurwitz matrix.*

(a) $E^c(p) = pW_cp'$ is a quadratic form with the coefficient matrix

$$W_c = \int_0^{\infty} e^{At} BB' e^{A't} dt.$$

(b) W_c is the unique solution of the Lyapunov equation

$$AW_c + W_c A' = -BB'. \quad (9.11)$$

(c) A given state $x_0 \in \mathbf{R}^n$ is reachable from zero if and only if $E_c(x_0) > 0$ or, equivalently, the equation $W_c p' = x_0$ has a solution p' . In this case $E_c(x_0) = p x_0$ is the minimum of $\|f(\cdot)\|^2$ subject to $X(f(\cdot)) = x_0$.

Proof. To prove (a), note that

$$\max_{\|f\| \leq 1} \int_{-\infty}^{\infty} g(t)' f(t) dt = \|g\|,$$

hence

$$E^c(p) = \int_0^{\infty} |p e^{At} B|^2 dt = p W_c p'.$$

Statement (b) is actually a re-wording of Theorem 9.9, with C replaced by B' , A replaced by A' , W_o replaced by W_c , and $x(0)$ replaced by p .

To prove (c), consider first the case when equation $W_c p' = x_0$ has no solution p . Then there exists a row vector p_0 such that $p_0 W_c = 0$ but $p_0 x_0 \neq 0$. Here the equality means that $|p_0 X(f(\cdot))|^2$ equals zero for every finite energy signal $f = f(t)$. Since, from the inequality, $|p_0 x_0|^2 > 0$, the state x_0 is not reachable from zero.

Now assume that $x_0 = W_c p'$ for some p . Then $\|f\|^2 \geq p W_c p' = p x_0$ whenever $x_0 = X(f(\cdot))$. On the other hand, for

$$f(t) = \begin{cases} B' e^{-A't} p', & t \leq 0, \\ 0, & t > 0, \end{cases}$$

we have $\|f\|^2 = p x_0$ and $x_0 = X(f(\cdot))$. ■

When the pair (A, B) is controllable, and, hence, $W_c > 0$, the primal controllability measure $E_c = E_c(x_0)$ can be expressed as

$$E_c(x_0) = \frac{1}{x_0' W_c^{-1} x_0} \quad \text{for } x_0 \neq 0.$$

9.3.4 Joint controllability and observability measures

The joint controllability and observability measures are defined as *products* of the corresponding controllability and observability measures:

$$E_{oc}(x_0) = E_o(x_0) E_c(x_0), \quad E^{oc}(p) = E^o(p) E^c(p).$$

For controllable and observable systems W_c and W_o are positive definite, and the formulae can be simplified to

$$E_{oc}(x_0) = \frac{x_0' W_o x_0}{x_0' W_c^{-1} x_0} \quad (x_0 \neq 0), \quad E^{oc}(p) = \frac{p W_c p'}{p W_o^{-1} p'} \quad (p \neq 0).$$

For model reduction purposes, we are interested in finding a subspace of primal state vectors for which the *minimum* of the joint controllability and observability measure over all non-zero elements is *maximal*. A basis in this subspace will yield columns of a projection matrix V . Similarly, we are interested in finding a subspace of dual state vectors for which the *minimum* of the joint controllability and observability measure over all non-zero elements is *maximal*. A basis in this subspace will yield rows of a projection matrix V .

The following theorem can be used in finding such V and U .

Theorem 9.11 *Let $W_c = L_c L'_c$ and $W_o = L'_o L_o$ be the Choleski decompositions of the controllability and observability Gramians. Let*

$$\rho_1 \geq \cdots \geq \rho_r > \rho_{r+1} \geq \cdots \geq \rho_n \geq 0$$

be the ordered eigenvalues of $L'_c W_o L_c$ (possibly repeated). Let $\psi_1, \dots, \psi_r$ be the corresponding first r normalized eigenvectors of $L'_c W_o L_c$, i.e.

$$L'_c W_o L_c \psi_i = \rho_i \psi_i, \quad |\psi_i|^2 = 1, \quad \psi'_i \psi_k = 0 \text{ for } i \neq k.$$

Let $\sigma_i = \rho_i^{1/2}$.

(a) $\rho_1 \geq \cdots \geq \rho_r$ are also the eigenvalues of $L_o W_c L'_o$, and the corresponding normalized row eigenvectors can be defined by

$$\phi_i = \sigma_i^{-1} \psi'_i L'_c L'_o \quad (i = 1, \dots, r).$$

(b) The set of all linear combinations of vectors $L_c \psi_i$ is the only r -dimensional linear subspace $\mathcal{V}$ in $\mathbf{R}^n$ such that $E_{oc}(v) \geq \rho_r$ for every $v \in \mathcal{V}$.

(c) The set of all linear combinations of row vectors $\phi_i L_o$ is the only r -dimensional linear subspace $\mathcal{U}$ of row n -vectors such that $E^{oc}(u) \geq \rho_k$ for every $u \in \mathcal{U}$.

(d) $UV = I_r$, where

$$V = L_c \begin{bmatrix} \psi_1 \sigma_1^{-1/2} & \cdots & \psi_r \sigma_r^{-1/2} \end{bmatrix}, \quad U = \begin{bmatrix} \sigma_1^{-1/2} \phi_1 \\ \vdots \\ \sigma_r^{-1/2} \phi_r \end{bmatrix} L_o.$$

The proof of the theorem is obtained by inspection.

9.3.5 Classical balanced truncation

The projection model reduction algorithm which uses the projection matrices U, V described in the previous subsection is called *balanced truncation*. The “balancing” terminology comes from the following trivial observation.

Theorem 9.12 *Assume that system (9.8) is both controllable and observable. Then $W_c = L_c L'_c > 0$ and $W_o = L'_o L_o > 0$ are positive definite. Moreover, if Ψ is the orthogonal matrix whose columns are the ordered eigenvectors of $L'_c W_o L_c$, and Σ is the diagonal matrix with numbers σ_i on the diagonal, then the linear state transformation $x = Sz$, with $S = L_c \Psi \Sigma^{-1/2}$, yields an equivalent state space model with the coefficient matrices*

$$\bar{A} = S^{-1}AS, \quad \bar{B} = S^{-1}B, \quad \bar{C} = CS,$$

for which both the controllability and observability Gramians equal Σ .

State space models for which $W_o = W_c$ are equal diagonal matrices with ordered positive diagonal entries σ_k are called *balanced*. The numbers σ_k are called the (Hankel) *singular numbers* of the corresponding system. The (canonical) method of balanced truncation is based on removing the states of a balanced realization which correspond to singular numbers below a certain threshold. Indeed, a practical implementation does not have to involve a calculation of a complete balanced model: only the projection matrices U, V are necessary.

Example Let

$$G(s) = \frac{1}{s+1-\epsilon} + \frac{1}{s+1+\epsilon},$$

where $\epsilon > 0$ is a small parameter. A state space model is

$$A = \begin{bmatrix} -1+\epsilon & 0 \\ 0 & -1-\epsilon \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad C = [1 \quad 1], \quad D = 0.$$

The controllability Gramians are given by

$$W_o = W_c = \frac{1}{2} \begin{bmatrix} \frac{1}{1-\epsilon} & 1 \\ 1 & \frac{1}{1+\epsilon} \end{bmatrix}.$$

The Hankel singular numbers are the eigenvalues of $W_o = W_c$, and equal

$$\sigma_{1,2} = \frac{1}{2} \frac{1 \pm \sqrt{1 - \epsilon^2 + \epsilon^4}}{1 - \epsilon^2}.$$

The corresponding eigenvectors are

$$v_{1,2} = \begin{bmatrix} 1 \\ 2\sigma_{1,2} - \frac{1}{1-\epsilon} \end{bmatrix}.$$

The dominant eigenvector defines

$$V \approx \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad U \approx \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The resulting reduced system is approximately given by

$$\hat{A} \approx -1, \quad \hat{B} \approx \sqrt{2}, \quad \hat{C} \approx \sqrt{2}.$$

9.3.6 Approximate Gramians

The two most expensive phases of numerical calculations associated with the canonical method of balanced truncation are finding the Gramians W_o, W_c , and calculating the dominant eigenvectors ψ_i of $L_i' W_o L_c$. for systems of large dimensions (more than 10^4 states) finding the Gramians exactly becomes difficult.

As a viable alternative, lower and upper bounds of the Gramians can be used to provide *provably* reliable results. Here by *lower bounds of Gramians* W_o, W_c we mean positive semidefinite matrices W_o^-, W_c^- for which the inequalities

$$W_o \geq W_o^-, \quad W_c \geq W_c^-$$

are guaranteed. The definition of upper bounds will be more strict: by *upper bounds of Gramians* W_o, W_c defined by the Lyapunov equalities

$$W_o A + A' W_o = -C' C, \quad A W_c + W_c A' = -B B',$$

where A is a Hurwitz matrix, we mean solutions W_o^+, W_c^+ of the corresponding Lyapunov inequalities

$$W_o^+ A + A' W_o^+ \leq -C' C, \quad A W_c^+ + W_c^+ A' \leq -B B'.$$

These inequalities imply that $W_o^+ \geq W_o$ and $W_c^+ \geq W_c$, but the inverse implication is not always true.

The following simple observation can be used to produce lower bounds of the Gramians.

Theorem 9.13 *Let A be an n -by- n Hurwitz matrix. Let F be an n -by- m matrix. For $s \in \mathbf{C}$ with $\operatorname{Re}(s) > 0$ define*

$$a = a(s) = (sI_n - A)^{-1}(\bar{s}I_n + A), \quad b = b(s) = \sqrt{2\operatorname{Re}(s)}(sI_n - A)^{-1}B.$$

Then

- (a) *a is a Schur matrix (all eigenvalues strictly inside the unit disc);*
- (b) *an n -by- n matrix P is a solution of the “continuous time” Lyapunov equation*

$$AP + PA' = -BB'$$

if and only if it is a solution of the “discrete time” Lyapunov equation

$$P = aPa' + bb';$$

- (c) *the matrix P from (b) is the limit*

$$P = \lim_{k \rightarrow \infty} P_k,$$

where the symmetric matrices $P_0 \leq P_1 \leq P_2 \leq \dots \leq P$ are defined by

$$P_0 = 0, \quad P_{k+1} = a(s_k)P_k a(s_k)' + b(s_k)b(s_k)',$$

and $\{s_k\}$ is a sequence of complex numbers contained in a compact subset of the open right half plane.

The theorem reduces finding a lower bound of a solution of a Lyapunov equation to solving systems of linear equations (used to produce $b(s_k)$ and the products $a(s_k)P_k a(s_k)'$).

Calculation of an upper bound of a Gramian could be more tricky. One approach to finding such upper bounds relies on having a valid *energy function* for A available.

Indeed, assume that $Q = Q'$ satisfies

$$AQ + QA' < 0.$$

If W_c^- is a good lower bound of the controllability Gramian W_c , defined by

$$AW_c + W_c A' = -BB',$$

then

$$AW_c^- + W_c^- A' \approx -BB',$$

and hence

$$A(W_c^- + \epsilon Q) + (W_c^- + \epsilon Q)A' \leq -BB'$$

for some $\epsilon > 0$ (which will, hopefully, be small enough). Then W_c^- and $W_c^- + \epsilon Q$ are a lower and an upper bound for the controllability Gramian W_c .

9.3.7 Lower bounds for model reduction error

A major benefit of doing balanced truncation is given by a *lower bound* of the error of *arbitrary* reduced models (not only those produced via balanced truncation).

Theorem 9.14 *Let $W_o^- = F_o' F_o$ and $W_c^- = F_c' F_c$ be lower bounds of the observability and controllability Gramians W_o, W_c of a stable LTI model $G = G(s)$. Let*

$$\sigma_1^- \geq \sigma_2^- \geq \dots \geq 0$$

be the ordered singular numbers of $F_o F_c$. Then σ_k^- is a lower bound for the k -th Hankel singular number $\sigma_k = \sigma_k(G)$ of the system, and

$$\|G - \hat{G}\|_\infty \geq \sigma_k^-$$

for every system $\hat{G}$ of order less than k .

Proof. Let $\mathcal{Z}_k$ denote the subspace spanned by the k dominant eigenvectors of $F_c' W_o^- F_c$, i.e.

$$|F_o F_c z| \geq \sigma_k^- |z| \quad \forall z \in \mathcal{Z}_k.$$

Since $W_c \geq F_c F_c'$, every vector $F_c z$ lies in the range of W_c , and $q' F_c z \leq |z|^2$ whenever $F_c z = W_c q$. Hence every state $x(0) = F_c z$ can be reached from $x(-\infty) = 0$ using a minimal energy input $f = f_z(t)$ (depending linearly on z) of energy not exceeding $|z|^2$. On the other hand, every state $x(0) = F_c z$ with $z \in \mathcal{Z}_k$ will produce at least $|F_o F_c z|^2 \geq (\sigma_k^-)^2 |z|^2$ of output energy. Since $\hat{G}$ is a linear system of order less than k , there exists at least one non-zero $z \in \mathcal{Z}_k$ for which the input $f = f_z(t)$ produces a zero state $\hat{x}(0) = 0$ at zero time. Then, assuming the input is zero for $t > 0$, the error output energy is at least $(\sigma_k^-)^2 |z|^2$. Since the testing input energy is not larger than $|z|^2 > 0$, this yields an energy gain of $(\sigma_k^-)^2$, which means that $\|G - \hat{G}\|_\infty \geq \sigma_k^-$. ■

9.3.8 Upper bonds for balanced truncation errors

The result from the previous subsection states that, for a stable LTI system G , no method can produce a reduced model $\hat{G}$ of order less than k such that the H-Infinity error $\|G - \hat{G}\|_\infty$ is less than the k -th singular number $\sigma_k = \sigma_k(G)$. The statement is easy to apply, since lower bounds σ_k^- of σ_k can be computed by using lower bounds of system Gramians.

The following theorem gives an *upper bound* of model reduction error for the *exact* implementation of the balanced truncation method.

Theorem 9.15 *Let $\sigma_1 > \sigma_2 > \dots > \sigma_h$ be the ordered set of different Hankel singular numbers of a stable LTI system G . Let $\hat{G}$ be the reduced model obtained by removing the states corresponding to singular numbers not larger than σ_k from a balanced realization of G . Then $\hat{G}$ is stable, and satisfies*

$$\|G - \hat{G}\|_\infty \leq 2 \sum_{i=k}^h \sigma_i.$$

The utility of Theorem 9.15 in practical calculations of H-Infinity norms of model reduction errors is questionable: an *exact* calculation of the H-Infinity norm is possible at about the same cost, and the upper bound itself can be quite conservative. Nevertheless, the theorem provides an important reassuring insight into the potential of balanced truncation: since the singular numbers of exponentially stable LTI systems decay exponentially, the upper bound of Theorem 9.15 is not expected to be much larger than the lower bound.

For example, for a system with singular numbers $\sigma_i = 2^{-i}$, a k th order reduced model cannot have quality better than 2^{-k-1} , and exact balanced truncation is guaranteed to provide quality of at least 2^{-k+1} .

The proof of Theorem 9.15 is based on estimating the quality of balanced truncation in the case when only the states of a balanced realization corresponding to the smallest Hankel singular number are removed, which is done in the following technical lemma.

Lemma 9.1 *Let $W = W' > 0$ be a positive definite symmetric n -by- n matrix satisfying the Lyapunov equalities*

$$WA + A'W = -C'C, \quad AW + WA' = -BB'. \quad (9.12)$$

Assume that W, A, B, C can be partitioned as

$$W = \begin{bmatrix} P & 0 \\ 0 & \sigma I_r \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C' = \begin{bmatrix} C'_1 \\ C'_2 \end{bmatrix},$$

where A_{22} is an r -by- r matrix, and matrices B_2, C'_2 have r rows. Then

(a) transfer matrices

$$G(s) = C(sI_n - A)^{-1}B, \quad G_1(s) = C_1(sI_{n-r} - A_{11})^{-1}B_1$$

are stable;

(b) the Lyapunov equalities

$$PA_{11} + A'_{11}P = -C'_1C_1, \quad A_{11}P + PA'_{11} = -B_1B'_1$$

are satisfied;

(c) $\|G - G_1\|_\infty \leq 2\sigma.$

Proof. It is sufficient to consider the case when the dimension m of $f = f(t)$ equals the dimension k of $y = y(t)$. (If $m < k$, add zero columns to B , if $m > k$, add zero rows to C .) First note that re-writing (9.12) in terms of the blocks A_{ik}, B_i, C_i yields

$$PA_{11} + A'_{11}P = -C'_1C_1, \tag{9.13}$$

$$PA_{12} + \sigma A'_{21} = -C'_1C_2, \tag{9.14}$$

$$\sigma(A_{22} + A'_{22}) = -C'_2C_2, \tag{9.15}$$

$$A_{11}P + PA'_{11} = -B_1B'_1, \tag{9.16}$$

$$\sigma A_{12} + PA'_{21} = -B_1B'_2, \tag{9.17}$$

$$\sigma(A_{22} + A'_{22}) = -B_2B'_2. \tag{9.18}$$

Note (9.15) together with (9.18) implies that $C'_2 = B_2\theta$ for some unitary matrix θ . Also, (9.13) and (9.16) prove (b).

To prove (a), note that for every complex eigenvector $v \neq 0$ of A , $Av = sv$ for some $s \in \mathbf{C}$, multiplication of the first equation in (9.12) by v' on the left and by v on the right yields

$$2\operatorname{Re}(s)v'Wv = -|Cv|^2.$$

Hence either $\operatorname{Re}(s) < 0$ or $\operatorname{Re}(s) = 0$ and $Cv = 0$. Hence all unstable modes of A are unobservable, and $G = G(s)$ has no unstable poles. The same proof applies to G_1 , since A_{11} satisfies similar Lyapunov equations.

To prove (c), consider the following state space model of the error system $G - G_1$:

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + A_{12}x_2 + B_1f, \\ \dot{x}_2 &= A_{21}x_1 + A_{22}x_2 + B_2f, \\ \dot{x}_3 &= A_{11}x_3 + B_1f, \\ e &= C_1x_1 + c_2x_2 - C_1x_3. \end{aligned}$$

It would be sufficient to find a positive definite quadratic form $V(x) = x'Hx$ such that

$$\psi(t) = 4\sigma^2|f(t)|^2 - |e(t)|^2 - \frac{dV(x(t))}{dt} \geq 0$$

for all solutions of system equations. Indeed, such Lyapunov function V can be readily presented, though there is no easy way to describe the intuitive meaning of its format:

$$V(x) = \sigma^2(x_1 + x_3)'P^{-1}(x_1 + x_3) + (x_1 - x_3)'P(x_1 - x_3) + 2\sigma|x_2|^2.$$

To streamline the derivation, introduce the shortcut notation

$$z = x_1 + x_3, \Delta = x_1 - x_3, \delta = C_1\Delta, u = \sigma^{-1}B_2'x_2, q = B_1'P^{-1}z.$$

The equations now take the form

$$\begin{aligned}\dot{\Delta} &= A_{11}\Delta + A_{12}x_2, \\ \dot{z} &= A_{11}z + A_{12}x_2 + 2B_1f, \\ \dot{x}_2 &= A_{22}x_2 + 0.5A_{21}(z + \Delta) + B_2f, \\ e &= C_1\Delta + \sigma\theta'u.\end{aligned}$$

We have

$$\begin{aligned}\psi &= 4\sigma^2|f|^2 - |C_1\Delta + \sigma\theta'u|^2 - 2\sigma^2z'P^{-1}(A_{11}z + A_{12}x_2 + 2B_1f) \\ &\quad - 2\Delta'P(A_{11}\Delta + A_{12}x_2) - 4\sigma x_2'[A_{22}x_2 + 0.5A_{21}(z + \Delta) + B_2f] \\ &= 4\sigma^2|f|^2 - |C_1\Delta + \sigma\theta'u|^2 + \sigma^2|q|^2 - 4\sigma^2q'f + \\ &\quad |\delta|^2 + 2\sigma^2|u|^2 - 4\sigma^2u'f - 2z'[\sigma^2P^{-1}A_{12} + \sigma A_{21}']x_2 - 2\Delta'[PA_{12} + \sigma A_{21}']x_2 \\ &= 4\sigma^2|f|^2 - |C_1\Delta + \sigma\theta'u|^2 + \sigma^2|q|^2 - 4\sigma^2q'f + |\delta|^2 \\ &\quad + 2\sigma^2|u|^2 - 4\sigma^2u'f + 2\sigma^2q'u + 2\sigma\delta'\theta'u \\ &\geq 4\sigma^2|f|^2 - |C_1\Delta + \sigma\theta'u|^2 + |\delta|^2 + 2\sigma^2|u|^2 + 2\sigma\delta'\theta'u - \sigma^2|u - 2f|^2 \\ &= 0\end{aligned}$$

(the first identity transformation utilized (9.13), (9.16), (9.15), the second identity used (9.14), (9.17), the third (inequality) applied minimization with respect to q , the last (identity) depended on θ being a unitary matrix. ■

It is important to realize that the lemma remains valid when the original Lyapunov equalities are replaced by the corresponding Lyapunov inequalities (of course, the equalities in (b) will get replaced by inequalities as well). Indeed, Lyapunov inequalities

$$WA + A'W \leq -C'C, \quad AW + WA' \leq -BB'$$

are equivalent to Lyapunov equalities

$$WA + A'W = -\tilde{C}'\tilde{C}, \quad AW + WA' = -\tilde{B}\tilde{B}',$$

where

$$\tilde{B} = \begin{bmatrix} B & B_\delta \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} C \\ C_\delta \end{bmatrix},$$

and B_δ, C_δ are chosen appropriately. Note that $G(s)$ is a left upper corner block in the transfer matrix

$$\tilde{G} = \tilde{C}(sI_n - A)^{-1}\tilde{B}.$$

Since H-Infinity norm of a transfer matrix is not smaller than H-Infinity norm of its block, applying Lemma 9.1 to the system defined by $A, \tilde{B}, \tilde{C}$ yields the stated generalization.

9.4 Hankel Optimal Model Reduction

Compared to the classical balanced truncation, Hankel optimal model reduction provides better model reduction errors at the same computational cost. Practically, however, the Hankel optimal model reduction algorithm is less flexible and more sensitive to numerical errors.

9.4.1 Hankel operators

Let L_r^2 denote the set of all integrable functions $e : \mathbf{R} \mapsto \mathbf{R}^r$ such that

$$\int_{-\infty}^{\infty} |e(t)|^2 dt < \infty.$$

Let $L_r^2(-\infty, 0)$ denote the subset of L_r^2 which consist of functions e such that $e(t) = 0$ for $t \geq 0$. The elements of $L_r^2(-\infty, 0)$ will be called *anti-causal* in this lecture. Similarly, let $L_r^2(0, \infty)$ be the subset of functions $e \in L_r^2$ such that $e(t) = 0$ for $t < 0$. The elements of $L_r^2(0, \infty)$ will be called *causal*.

Let $G = G(s)$ be a k -by- m matrix-valued function (not necessarily a rational one), bounded on the $j\omega$ -axis. The corresponding *Hankel operator* $\mathcal{H} = \mathcal{H}_G$ is the linear transformation which maps anti-causal square integrable functions $f \in L_r^2(-\infty, 0)$ to causal square integrable functions $h = \mathcal{H}_G f \in L_r^2(0, \infty)$ according to the following rule: $h(t) = y(t)u(t)$, where $y(t)$ is the inverse Fourier transform of $Y(j\omega) = G(j\omega)F(j\omega)$, $F(j\omega)$ is the Fourier transform of $f(t)$, and $u(t)$ is the unit step function

$$u(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

In terms of the (stable, but not necessarily causal) LTI system defined by G , the Hankel operator maps anti-causal inputs $f = f(t)$ to the causal parts $h = h(t) = y(t)u(t)$

of the complete system response $y(t)$. In particular, when G is anti-stable, i.e. is a proper rational transfer matrix without poles s with $\text{Re}(s) \leq 0$, the associated LTI system is anti-causal, and hence the resulting Hankel operator $\mathcal{H}_G$ is zero. More generally, adding an anti-stable component to G does not affect the resulting $\mathcal{H}_G$.

9.4.2 Hankel matrices

Let $a > 0$ be a fixed positive number. Then functions

$$\Theta_k(j\omega) = \frac{\sqrt{2a}}{s+a} \left(\frac{a-s}{a+s} \right)^k, \quad k = 0, 1, 2, \dots$$

form an orthonormal basis in the space of stable strictly proper transfer functions, in the sense that for every such function $H = H(s)$ there exists a square summable sequence of real numbers $h_0, h_1, h_2, \dots$ satisfying

$$H(j\omega) = \sum_{k=0}^{\infty} h_k \Theta_k(j\omega),$$

i.e.

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H(j\omega) - \sum_{k=0}^N h_k \Theta_k(j\omega) \right|^2 d\omega = \sum_{k=N+1}^{\infty} |h_k|^2 \rightarrow 0 \text{ as } N \rightarrow \infty.$$

In a similar sense, the inverse Fourier transforms $\theta_k = \theta_k(t)$ of $\Theta_k = \Theta_k(j\omega)$, form an orthonormal basis in $L_1^2(0, \infty)$, and the inverse Fourier transforms $\theta_k(-t)$ of $\Theta_k(-j\omega)$ form an orthonormal basis in $L_1^2(0, \infty)$.

The following lemma allows one to establish a matrix representation of a Hankel operator with respect to input basis $\{\theta_k(-t)\}_{k=0}^{\infty}$ and output basis $\{\theta_k(t)\}_{k=0}^{\infty}$.

Lemma 9.2 *Let*

$$g_k = \frac{1}{\pi} \int_{-\infty}^{\infty} G(j\omega) \left(\frac{a+j\omega}{a-j\omega} \right)^k \frac{a d\omega}{a^2 + \omega^2}.$$

Then the result $h = h(t)$ of applying $\mathcal{H}_G$ to $f(t) = \theta_r(-t)$ is given by

$$h(t) = \sum_{k=0}^{\infty} g_{r+k+1} \theta_k(t).$$

An important implication of the lemma is that the matrix of $\mathcal{H}_G$ with respect to the input/output bases $\{\theta_k(-t)\}_{k=0}^\infty, \{\theta_k(t)\}_{k=0}^\infty$ is the *Hankel matrix*

$$\Gamma_G = \begin{bmatrix} g_1 & g_2 & g_3 & g_4 & & \\ g_2 & g_3 & g_4 & & & \\ g_3 & g_4 & & & & \\ g_4 & & & & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}.$$

In general, g_k are matrices with real coefficients, in which case Γ_G is called the *block Hankel matrix*.

Proof. Consider the decomposition

$$h(t) = \sum_{k=0}^{\infty} h_k \theta_k(t).$$

By orthonormality of $\theta_k(\cdot)$,

$$h_k = \int_0^\infty h(t) \theta_k(t) dt = \int_{-\infty}^\infty y(t) \theta_k(t) dt,$$

where y is the response of the stable LTI system associated with G to $f(t) = \theta_r(-t)$. By the Parseval formula,

$$\begin{aligned} h_k &= \frac{1}{2\pi} \int_{-\infty}^\infty \Theta_k(-j\omega) G(j\omega) \Theta_r(-j\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^\infty G(j\omega) \left(\frac{a+j\omega}{a-j\omega} \right)^{k+r+1} \frac{2ad\omega}{(a+j\omega)(a-j\omega)} = g_{k+r+1}. \end{aligned}$$

■

9.4.3 Singular values of a Hankel operator

Let M be an a -by- b matrix representing a linear transformation from $\mathbf{R}^b$ to $\mathbf{R}^a$. Remember that the *operator norm* of M is defined as the minimal upper bound for all ratios $|Mv|/|v|$, where v ranges over the set of all non-zero vectors in $\mathbf{R}^b$. In addition, the r -th singular number of M can be defined as the minimal operator norm of the difference $\Delta = M - \hat{M}$, where $\hat{M}$ ranges over the set of all matrices with rank less than r .

These definitions extend naturally to linear transformations of other normed vector spaces (possibly infinite dimensional). In particular, for a linear transformation M from

$L_m^2(-\infty)$ to $L_k^2(0, \infty)$, its *operator norm* is defined as the square root of the minimal upper bound for the ratio

$$\int_0^\infty |(Mf)(t)|^2 dt / \int_{-\infty}^0 |f(t)|^2 dt,$$

where

$$\int_{-\infty}^0 |f(t)|^2 dt > 0.$$

Such transformation M is said to have *rank* less than r if for every family of r functions $f_1, \dots, f_r \in L_k^2(-\infty, 0)$ there exist constants $c_1, \dots, c_r$, not all equal to zero, such that

$$c_1(Mf_1) + \dots + c_r(Mf_r) \equiv 0.$$

Finally, the r -th singular number of M can be defined as the minimal operator norm of the difference $\Delta = M - \hat{M}$, where $\hat{M}$ ranges over the set of all matrices with rank less than r .

This allows us to talk about the k -th singular number of the Hankel operator $\mathcal{H}_G$ associated with a given matrix-valued function $G = G(j\omega)$, bounded on the imaginary axis. The largest singular number is called the *Hankel norm* $\|G\|_H$ of G , while the k -th singular number is called the k -th *Hankel singular number* of G .

For rational transfer matrices G , calculation of singular numbers of the corresponding Hankel operator can be done using observability and controllability Gramians. The following theorem was, essentially, proven in the lectures on balanced truncation.

Theorem 9.16 *Let A be an n -by- n Hurwitz matrix, B, C be matrices of dimensions n -by- m and k -by- n respectively, such that the pair (A, B) is controllable, and the pair (C, A) is observable. Let W_c, W_o be the corresponding controllability and observability Gramians. Then, for $G(s) = C(sI - A)^{-1}B$, the Hankel operator $\mathcal{H}_G$ has exactly n positive singular numbers, which are the square roots of the eigenvalues of $W_c W_o$.*

It is also true that a Hankel operator with a finite number of positive singular numbers is defined by a rational transfer matrix.

Theorem 9.17 *Let $G = G(j\omega)$ be a bounded matrix-valued function defined on the imaginary axis. Let $a > 0$ be a positive number. If the Hankel operator $\mathcal{H}_G$ has less than r positive singular numbers then the coefficients*

$$g_k = \frac{1}{\pi} \int_{-\infty}^{\infty} G(j\omega) \left(\frac{a + j\omega}{a - j\omega} \right)^k \frac{a d\omega}{a^2 + \omega^2}$$

coincide for $k > 0$ with such coefficients of a stable strictly proper transfer matrix G_1 of order less than r .

For some non-rational transfer matrices, analytical calculation of σ_i may be possible. For example, the i -th largest singular number of $\mathcal{H}_G$, where $G(s) = \exp(-s)$, equals 1 for all positive i .

In general, singular numbers of $\mathcal{H}_G$ will converge to zero if $G = G(j\omega)$ is continuous on the extended imaginary axis (note that $G(s) = \exp(-s)$ is not continuous at $\omega = \infty$). The converse statement is not true.

The Hankel optimal model reduction setup Let $G = G(s)$ be a matrix-valued function bounded on the $j\omega$ -axis. The task of Hankel optimal model reduction of G calls for finding a stable LTI system $\hat{G}$ of order less than a given positive integer m , such that the Hankel norm $\|\Delta\|_H$ of the difference $\Delta = G - \hat{G}$ is minimal.

Since Hankel operator $\mathcal{H}_G$ represents a “part” of the total LTI system with transfer matrix G , Hankel norm is never larger than H-Infinity norm. Hence, Hankel optimal model reduction setup can be viewed as a *relaxation* of the “original” (H-Infinity optimal) model reduction formulation. While no acceptable solution is available for the H-infinity case, Hankel optimal model reduction has an elegant and algorithmically efficient solution.

9.4.4 The AAK theorem

The solution of the Hankel optimal model reduction problem is based on the famous Adamyan-Arov-Krein (AAK) theorem, which provides both a theoretical insight and (taking a constructive proof into account) an explicit algorithm for finding Hankel optimal reduced models.

Theorem 9.18 *Let $G = G(s)$ be a matrix-valued function bounded on the $j\omega$ -axis. Let $\sigma_1 \geq \sigma_2 \geq \dots \sigma_m \geq 0$ be the m largest singular values of $\mathcal{H}_G$. Then σ_m is the minimum of $\|G - \hat{G}\|_H$ over the set of all stable systems $\hat{G}$ of order less than m .*

In other words, approximating Hankel operators by general linear transformations of rank less than m cannot be done better (in terms of the minimal L2 gain of the error) than approximating it by Hankel operators of rank less than m .

The proof of the theorem, to be given in this section for the case of a rational transfer matrix $G = G(s)$, is constructive, and provides a simple state space algorithm for calculating the Hankel optimal reduced model.

H-Infinity quality of Hankel optimal reduced models It is well established by numerical experiments that Hankel optimal reduced models usually offer very high H-Infinity quality of model reduction. A somewhat conservative description of this effect is given by the following extension of the AAK theorem.

Theorem 9.19 *Let $G = G(s)$ be a stable rational function. Assume that the Hankel singular numbers $\sigma_k = \sigma_k(G)$ of G satisfy*

$$\sigma_{m-1} > \sigma_m = \sigma_{m+1} = \cdots = \sigma_{m+r-1} > \sigma_{m+r}.$$

$\sigma_k(G) = \sigma_m(G)$ for $m \leq k < m+r$, and $\sigma_{m+r}(G) < \sigma_m(G)$. Let $\tilde{\sigma}_m > \tilde{\sigma}_{m+1} > \tilde{\sigma}_{m+2} > \cdots$ be the ordered sequence of different Hankel singular values of G , starting with $\tilde{\sigma}_m = \sigma_m$ and $\tilde{\sigma}_{m+1} = \sigma_{m+r}$. Then

(a) *there exists a Hankel optimal reduced model $\hat{G}_m^H$ of order less than m such that*

$$\|G - \hat{G}_m^H\|_\infty \leq \sigma_m + \sum_{k>m}^r \sigma_k;$$

(b) *there exists a model $\hat{G}_m^*$ of order less than m such that*

$$\|G - \hat{G}_m^*\|_\infty \leq \sum_{k \geq m} \tilde{\sigma}_k.$$

Just as in the case of the basic AAK theorem, the proof of Theorem 9.19 is constructive, and hence provides an explicit algorithm for calculation of reduced models with the described properties. In practice, the actual H-Infinity norm of model reduction error is much smaller.

It is important to remember that the Hankel optimal reduced model is never unique (at least, the “D” terms do not have any effect on the Hankel norm, and hence can be modified arbitrarily). The proven H-Infinity model reduction error bound is guaranteed only for a specially selected Hankel optimal reduced model. Also, the reduced model from (b) is not necessarily a Hankel optimal reduced model.

AAK theorem: general comments on the proof It is sufficient to consider the case when the dimension of $f = f(t)$ equals the dimension of $y = y(t)$ (otherwise, add zero columns to B or zero rows to C).

Since Hankel operator of an anti-stable system is zero, and rank of a Hankel operator of a system of order less than m is less than m , the inequality

$$\|G - \hat{G}_m\|_H \geq \sigma_m(G)$$

holds when the order of $\hat{G}_m$ is less than m .

What remains to be proven is the existence of a $\hat{G}_m$ of order less than m such that

$$\|G - \hat{G}_m\|_H \leq \sigma_m(G).$$

This will be done by constructing explicitly a state space model of transfer matrix $L(s) = \hat{G}_m^H(s) + F^H(s)$, where $\hat{G}_m^H$ is stable and has order less than m , F^H is anti-stable, and $\|G - L\|_\infty = \sigma_m(G)$. Then, by definition, $\|G - \hat{G}_m^H\|_H \leq \sigma_m(G)$.

Actually, a stronger conclusion will be reached: L can be chosen in such way that

$$E(j\omega)'E(j\omega) = \sigma_m(G)I \quad \forall \omega \in \mathbf{R} \quad (9.19)$$

for $E = H - L$. Condition (9.19) will be used later to derive upper bounds for $\|G - \hat{G}_m^H\|_\infty$.

Partitions of the coefficient matrices Assume that G is defined by a minimal (controllable and observable) *balanced* finite dimensional state space model

$$\dot{x} = Ax + Bf, \quad y = Cx \quad (9.20)$$

with a Hurwitz n -by- n matrix A . Without loss of generality, consider the case when the controllability and observability Gramian $W = W_c = W_o$ of (9.20) has the form

$$W = \begin{bmatrix} \Sigma & 0 \\ 0 & \gamma I_r \end{bmatrix},$$

where $\gamma = \sigma_m(G)$, the m -th singular number of G (multiplicity r), is *not* an eigenvalue of $\Sigma = \Sigma' > 0$.

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \quad C_2]$$

be the corresponding block partitions of A, B, C (for example, A_{22} is an r -by- r matrix). Since

$$AW + WA' = -BB', \quad WA + A'W = -C'C,$$

the blocks A_{ij} , B_i , C_i satisfy the relations

$$\Sigma A_{11} + A'_{11} \Sigma = -C'_1 C_1, \quad (9.21)$$

$$\Sigma A_{12} + \gamma A'_{21} = -C'_1 C_2, \quad (9.22)$$

$$\gamma(A_{22} + A'_{22}) = -C'_2 C_2, \quad (9.23)$$

$$A_{11} \Sigma + \Sigma A'_{11} = -B_1 B'_1, \quad (9.24)$$

$$\gamma A_{12} + \Sigma A'_{21} = -B_1 B'_2, \quad (9.25)$$

$$\gamma(A_{22} + A'_{22}) = -B_2 B'_2. \quad (9.26)$$

Let

$$\Delta = \Sigma - \gamma^2 \Sigma^{-1}.$$

Combining (9.13) with (9.16) (multiplied by $\gamma \Sigma^{-1}$ on both sides) yields

$$\Delta A_{11} + A_{11} \Delta = \gamma^2 \Sigma^{-1} B_1 B'_1 \Sigma^{-1} - C'_1 C_1. \quad (9.27)$$

Similarly, combining (9.14) with (9.17) yields

$$A'_{12} \Delta = -C'_2 C_1 + \gamma B_2 B'_1 \Sigma^{-1}. \quad (9.28)$$

Finally, (9.15) together with (9.18) implies that $B_2 B'_2 = C'_2 C_2$, which in turn means that $C'_2 U = B_2$ for some unitary matrix U .

In the following section, it will be useful to know that A_{11} has no eigenvalues on the imaginary axis (though in general it may have eigenvalues with positive and negative real parts).

Lemma 9.3 *Let*

$$a = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad c = \begin{bmatrix} c_1 & c_2 \end{bmatrix}, \quad p = p' = \begin{bmatrix} q & 0 \\ 0 & \gamma I \end{bmatrix}$$

be such that

$$pa + a'p = -c'c, \quad ap + pa' = -bb'.$$

If γ^2 is not an eigenvalue of q and a has no eigenvalues on the imaginary axis then a_{11} has no eigenvalues on the imaginary axis.

Proof. Assume to the contrary that $a_{11}f = j\omega f$ for some $f \neq 0$. Then

$$-|c_1 f|^2 = -f' c'_1 c_1 f = f'(q a_{11} + a'_{11} q) f = j\omega f' q f - j\omega f' q f = 0.$$

Hence $c_1 f = 0$.

$$a'_{11} q f = (-c'_1 c_1 - q a_{11}) f = -j\omega q f.$$

Then

$$-|b'_1 q f|^2 = -(q f)' b_1 b'_1 (q f) = (q f)' (a_{11} q + q a'_{11}) (q f) = 0.$$

Hence $b'_1 q f = 0$. Therefore

$$a_{11} q^2 f = (-b_1 b'_1 - q a'_{11}) q f = j\omega q^2 f,$$

which, combined with $a_{11} f = j\omega f$, yields

$$a_{11}(q^2 - \gamma^2 I) f = j\omega(q^2 - \gamma^2 I) f. \quad (9.29)$$

On the other hand, equalities

$$a'_{12} q + \gamma a_{21} = -c'_2 c_1, \quad a_{21} q + \gamma a'_{12} = -b_2 b'_1$$

imply

$$a_{21}(q^2 - \gamma^2) f = 0. \quad (9.30)$$

Combining (9.29) and (9.30) yields

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} (q^2 - \gamma^2) f \\ 0 \end{bmatrix} = j\omega \begin{bmatrix} (q^2 - \gamma^2) f \\ 0 \end{bmatrix},$$

which contradicts the assumptions. ■

9.4.5 AAK theorem proof: explicit formulae and certificates

In terms of the matrices introduced in the previous subsection, it is easy to define explicitly a state space model of L , as well as the certificates of the H-Infinity norm bounds for $G-L$.

Lemma 9.4 *Let*

$$A_L = A_{11} - \Delta^{-1}(\gamma^2 \Sigma^{-1} B_1 - \gamma C'_1 U) B'_1 \Sigma^{-1}, \quad (9.31)$$

$$B_L = B_1 + \Delta^{-1}(\gamma^2 \Sigma^{-1} B_1 - \gamma C'_1 U), \quad (9.32)$$

$$C_L = C_1 - \gamma U B'_1 \Sigma^{-1}, \quad (9.33)$$

$$D_L = \gamma U. \quad (9.34)$$

Then

(a) *the pair (A_L, B_L) is controllable;*

- (b) the pair (C_L, A_L) is observable;
- (c) A_L (dimension $n-r$) has $m-1$ eigenvalues with negative real part and $n-r-m+1$ eigenvalues with positive real part;
- (d) transfer matrix $E(s) = \gamma^{-1}(G(s) - L(s))$ satisfies

$$E(j\omega)'E(j\omega) = \gamma^2 I.$$

- (e) The identity

$$\gamma^2 |f|^2 - |Cx - C_L x_L - D_L f|^2 - 2\operatorname{Re} \left\{ \begin{bmatrix} x \\ x_L \end{bmatrix}' H \begin{bmatrix} Ax + Bf \\ A_L x_L + B_L f \end{bmatrix} \right\} = 0 \quad \forall x, f, x_L \quad (9.35)$$

holds for

$$H = H' = \begin{bmatrix} \Sigma & 0 & -\Delta \\ 0 & \gamma I_r & 0 \\ -\Delta & 0 & \Delta \end{bmatrix}.$$

Proof. Identity (9.35) in (e) can be checked “by inspection”.

Statement (d) follows from (e) by substituting x, x_L, f such that

$$j\omega x = Ax + Bf, \quad j\omega x_L = A_L x_L + B_L f,$$

since the real part equals zero in this case.

To prove (a),(b), note first that, according to (9.35), identities

$$Ha + a'H = -c'c, \quad aH^{-1} + H^{-1}a' = -bb',$$

where

$$H^{-1} = \begin{bmatrix} \gamma^{-2}\Sigma & 0 & \gamma^{-2}\Sigma \\ 0 & \gamma^{-1}I & 0 \\ \gamma^{-2}\Sigma & 0 & \gamma^{-2}\Sigma\Delta^{-1}\Sigma \end{bmatrix},$$

hold for

$$a = \begin{bmatrix} A & 0 \\ 0 & A_L \end{bmatrix}, \quad b = \begin{bmatrix} B \\ B_L \end{bmatrix}, \quad c = \begin{bmatrix} C & -C_L \end{bmatrix}.$$

Hence

$$\Delta A_L + A_L' \Delta = -C_L' C_L, \quad A_L(\Sigma\Delta^{-1}\Sigma) + (\Sigma\Delta^{-1}\Sigma)A_L' = -B_L B_L'.$$

Since Δ and $\Sigma\Delta^{-1}\Sigma$ are not singular, controllability of (A_L, B_L) and observability of (C_L, A_L) will follow if A_L has no eigenvalues on the imaginary axis. However, if $f'A_L = j\omega f'$ for some $f \neq 0$ then $f'B'_L = 0$. Since

$$A_L + B_L B'_1 \Sigma^{-1} = A_{11} + B_1 B'_1 \Sigma^{-1},$$

this would imply

$$f'(A_{11} + B_1 B'_1 \Sigma^{-1}) = j\omega f'.$$

Since

$$(A_{11} + B_1 B'_1 \Sigma^{-1})\Sigma + \Sigma(A_{11} + B_1 B'_1 \Sigma^{-1})' = B_1 B'_1,$$

this implies $f'B_1 = 0$ and hence $f'A_{11} = j\omega f'$, which is impossible due to Lemma 9.3.

Finally, since (A_L, B_L) is controllable, (C_L, A_L) is observable, and A_L has no eigenvalues on the imaginary axis, statement (c) follows. ■

9.4.6 KYP lemma for L-Infinity norm approximation

A simple but important observation given by the KYP lemma is that a “certificate” of an L-Infinity bound $\|H\|_\infty \leq \gamma$ for a given transfer matrix $H(s) = d + c(sI - a)^{-1}b$ is delivered by a symmetric matrix $p = p'$ such that

$$\gamma^2 |w|^2 - |c\bar{x} + dw|^2 - 2\bar{x}'p(a\bar{x} + bw) \geq 0 \quad \forall \bar{x}, w. \quad (9.36)$$

Note that a does not have to be a Hurwitz matrix here.

When system (9.20) is approximated by system G_r (not necessarily stable) with state space model

$$\dot{x}_r = A_r x_r + B_r w, \quad y_r = C_r x_r + D_r w, \quad (9.37)$$

the “approximation error dynamics” system with input w and output $\delta = y - y_r$ has state space model

$$\dot{\bar{x}} = a\bar{x} + bw, \quad \delta = c\bar{x} + dw,$$

where

$$\bar{x} = \begin{bmatrix} x \\ x_r \end{bmatrix}, \quad a\bar{x} + bw = \begin{bmatrix} Ax + Bw \\ A_r x_r + B_r w \end{bmatrix}, \quad c\bar{x} + dw = Cx + Dw - C_r x_r - D_r w. \quad (9.38)$$

Let $\Delta = \Delta(s)$ denote the transfer matrix from w to δ . According to the KYP lemma, the inequality $\|\Delta\|_\infty \leq \gamma$ can be established by finding a symmetric matrix

$$p = p' = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad (9.39)$$

such that (9.36) holds.

For Hankel model order reduction it is important to keep track of the order of the stable part of G_r . This is made possible by the following observation.

Theorem 9.20 *If (9.36) holds for a, b, c, d, p defined by (9.38), (9.39), and p_{22} has less than m positive eigenvalues then the order of the stable part of G_r is less than m .*

Proof. Let $V = V_+ \dot{+} V_-$ be the direct sum decomposition of the state space $\{x_r\}$ into the *stable observable* subspace V_+ of A_r with respect to C_r , and the complementary A_r -invariant subspace V_- . In a system of coordinates associated with this decomposition, matrices C_r, A_r, p_{22} have the block form

$$C_r = \begin{bmatrix} c_+ & c_- \end{bmatrix}, \quad A_r = \begin{bmatrix} a_+ & 0 \\ 0 & a_- \end{bmatrix}, \quad p_{22} = \begin{bmatrix} p_{++} & p_{+-} \\ p_{-+} & p_{--} \end{bmatrix},$$

where the pair (c_+, a_+) is observable. Substituting $w = 0$ into (9.36) yields

$$pa + a'p \leq -c'c,$$

which, in particular, implies

$$p_{++}a_+ + a'_+p_{++} \leq -c'_+c_+.$$

Hence $p_{++} > 0$ and the number of positive eigenvalues of p_{22} is at least as large as the dimension of a_+ . Finally, note that the dimension of a_+ is not smaller than the order of the stable part of G_r . ■

9.4.7 Hankel optimal reduced models via Parrot's Theorem

The observations made in the previous two subsections suggest the following approach to constructing a reduced model $\hat{G}$ of system G given by (9.20): simply find a matrix $p = p'$ in (9.39) such that p_{22} has less than $n + m$ positive eigenvalues, and (9.36) holds for a, b, c, d defined by (9.38) with some A_r, B_r, C_r, D_r . Then $\hat{G}$, defined as the stable part of $D_r + C_r(sI - A_r)^{-1}B_r$, will satisfy $\|G - \hat{G}\|_H \leq \gamma$.

Note that, once p is fixed, the existence of A_r, B_r, C_r, D_r satisfying the requirements can be checked via the generalized Parrot's theorem (same as used in the derivation of H-Infinity suboptimal controllers).

Theorem 9.21 *Let $\sigma : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^k \rightarrow \mathbf{R}$ be a quadratic form which is concave with respect to its second argument, i.e.*

$$\sigma(0, g, 0) \leq 0 \quad \forall g \in \mathbf{R}^m. \tag{9.40}$$

Then an m -by- k matrix L such that

$$\sigma(f, Lh, h) \geq 0 \quad \forall f \in \mathbf{R}^n, h \in \mathbf{R}^k \quad (9.41)$$

exists if and only if the following two conditions are satisfied:

(a) for every $h \in \mathbf{R}^k$ there exists $g \in \mathbf{R}^m$ such that

$$\inf_{f \in \mathbf{R}^n} \sigma(f, g, h) > -\infty; \quad (9.42)$$

(b) the inequality

$$\sup_{g \in \mathbf{R}^m} \sigma(f, g, h) \geq 0 \quad (9.43)$$

holds for all $f \in \mathbf{R}^n, h \in \mathbf{R}^k$.

In application to Hankel optimal model reduction, set

$$f = x, \quad g = \begin{bmatrix} \theta \\ y_r \end{bmatrix}, \quad h = \begin{bmatrix} x_r \\ w \end{bmatrix}, \quad L = \begin{bmatrix} A_r & B_r \\ C_r & D_r \end{bmatrix},$$

$$\sigma(f, g, h) = -2 \begin{bmatrix} x \\ x_r \end{bmatrix}' p \begin{bmatrix} Ax + Bw \\ \theta \end{bmatrix} + \gamma^2 |w|^2 - |Cx + Dw - y_r|^2. \quad (9.44)$$

Note that σ is concave with respect to g .

It turns out that one convenient selection for p is

$$p = \begin{bmatrix} W_o & W_o - \gamma^2 W_c^{-1} \\ W_o - \gamma^2 W_c^{-1} & W_o - \gamma^2 W_c^{-1} \end{bmatrix}. \quad (9.45)$$

While formally there is no need to explain this choice (one just has to verify that conditions (a),(b) of Theorem 9.21 are satisfied), there is a clear line of reasoning here. To come to this particular choice of p , note first that (9.42) implies $\sigma(f, 0, 0) \geq 0$, which means

$$p_{11}A + A'p_{11} \leq -C'C$$

for σ defined by (9.44). Hence

$$p_{11} \geq W_o.$$

Similarly, under the simplifying assumption that p is invertible, (9.43) means

$$Aq_{11} + q_{11}A' \leq -\gamma^{-2}BB',$$

where q_{11} is the upper left corner of p^{-1} . Hence

$$q_{11} \geq \gamma^{-2}W_c.$$

Since

$$q_{11}^{-1} = p_{11} - p_{12}p_{22}^{-1}p_{21},$$

we have

$$p_{12}p_{22}^{-1}p_{21} = p_{11} - \gamma^2q_{11}^{-1}.$$

Since our desire is to minimize the number of positive eigenvalues of p_{22} , it is natural to use the minimal possible values of p_{11} and q_{11} , which suggests using

$$p_{11} = W_o, \quad q_{11} = \gamma^{-2}W_c, \quad p_{12} = p_{21} = p_{22} = W_o - \gamma^2W_c^{-1}.$$

With the proposed selection of p , the quadratic form σ can be represented as

$$\sigma(f, g, h) = -2x'[W_oBw + A'(W_o - \gamma^2W_c^{-1})x_r + (W_o - \gamma^2W_c^{-1})\theta - C'y_r] + \tilde{\sigma}(g, h).$$

Hence condition (a) can be satisfied if the equation

$$W_oBw + A'(W_o - \gamma^2W_c^{-1})x_r + (W_o - \gamma^2W_c^{-1})\theta - C'y_r = 0$$

has a solution (θ, y_r) for every pair (w, x_r) . In order to prove this, it is sufficient to show that every vector ψ such that

$$\psi'C' = 0, \quad \psi'(W_o - \gamma^2W_c^{-1}) = 0 \tag{9.46}$$

also satisfies

$$\psi'W_oB = 0, \quad \psi'A'(W_o - \gamma^2W_c^{-1}) = 0. \tag{9.47}$$

Note that by the definition,

$$(W_o - \gamma^2W_c^{-1})A + A'(W_o - \gamma^2W_c^{-1}) + C'C = \gamma^2W_c^{-1}BB'W_c^{-1}.$$

Multiplying this by ψ' on the left and ψ on the right and using (9.46) yields $\psi'B'W_c^{-1} = 0$ and hence (9.47) follows.

Similarly, the quadratic form σ can be represented as

$$\sigma(f, g, h) = (x+x_r)'(W_o - \gamma^2W_c^{-1})(Ax+Bw-\theta) + 2\gamma^2x'W_c^{-1}(Ax+Bw) + \gamma^2|w|^2 - |Cx-y_r|^2,$$

which is unbounded with respect to θ if $(x+x_r)'(W_o - \gamma^2W_c^{-1}) \neq 0$ and is made non-negative by selecting $y_r = -Cx_r$ otherwise.

10 Convex optimization

Many optimization objectives generated by LTI system design and analysis do not fit within the frameworks of H2/H-Infinity optimization or Hankel optimal model reduction, but are still relatively easy to work with. In most cases, such objectives are characterized by *convexity* of the underlying constraints. This lecture covers the basic techniques associated with recognition and handling of convex optimization.

10.1 Basic theory of convex analysis

In this subsection, basic definitions of convex optimization are given.

10.1.1 Convexity

Recall that a set V is called a (real) *vector space* if it is equipped with operations of addition and multiplication by a real scalar, satisfying the usual conditions of distribution and commutation. The main examples of real vector spaces are $\mathbf{R}^n$ and $\mathbf{C}^n$. More generally, every set $V = \{f(\cdot)\}$ of functions $f : X \mapsto \mathbf{C}$, closed with respect to the natural operations of addition and scaling, is a vector space.

A subset Ω of a vector space $V = \mathbf{R}$ is called *convex* if

$$cv_1 + (1 - c)v_2 \in \Omega \text{ whenever } v_1, v_2 \in \Omega, c \in [0, 1].$$

In other words, a set is convex whenever the line segment connecting any two points of Ω lies completely within Ω . A function $\Phi : \Omega \mapsto \mathbf{R} \cup \{+\infty\}$ is called *convex* if Ω is a convex set, and the “overgraph”

$$\Gamma_\Phi = \{(v, y) \in \Omega \times \mathbf{R} : y > \Phi(v)\}$$

of Φ is convex, i.e. when

$$\Phi(cv_1 + (1 - c)v_2) \leq c\Phi(v_1) + (1 - c)\Phi(v_2) \text{ whenever } v_1, v_2 \in \Omega, c \in [0, 1].$$

A function $\Phi : \Omega \mapsto \mathbf{R} \cup \{+\infty\}$ is called *quasi-convex* if Ω is a convex set, and, for every $\gamma \in \mathbf{R}$, the set

$$\Omega_\gamma = \{v \in \Omega : \Phi(v) < \gamma\}$$

is convex.

The tasks of minimizing a convex or a quasi-convex function, or of finding an element of an implicitly defined convex set, are frequently referred to as *convex optimization*. As a rule, one expects a convex optimization problem to be “easy to solve”, though there are several other factors involved, such as dimension of the decision vector and complexity of description of Ω and/or Φ .

10.1.2 Dual representations of convex sets and functions

In many cases, it is inconvenient to apply the direct definitions of convexity, and a *duality* approach is used. Roughly speaking, the duality in convex optimization is based on defining convex sets as intersections of families (possibly infinite) of half-spaces. Similarly, the duality approach defines convex functions as point-wise maximums (strictly speaking, supremums) of families of affine functions.

Let us call a real-valued function $f : V \mapsto \mathbf{R}$ defined on a vector space V *linear functional* if

$$f(c_1v_1 + c_2v_2) = c_1f(v_1) + c_2f(v_2) \quad \text{whenever } v_1, v_2 \in V, c_1, c_2 \in \mathbf{R}.$$

An *affine functional* is a function $g : V \mapsto \mathbf{R}$ of the form $g(v) = f(v) + c$, where $f : V \mapsto \mathbf{R}$ is a linear functional, and c is a constant. An *open half-space* in V is a level set

$$V_g = \{v \in V : g(v) > 0\}$$

of an affine function $g : V \mapsto \mathbf{R}$.

The following statement follows directly from the definitions.

Lemma 10.1 *Let K be a set of affine functionals on vector space V . Then the subset Ω of V defined by*

$$\Omega_K = \{v \in V : g(v) > 0 \quad \forall g \in K\}$$

is convex. Moreover, for every convex subset $\Omega \subset V$ the function

$$\Phi(v) = \sup_{g \in K} g(v)$$

is convex on Ω .

In other word, a set defined by affine inequalities is convex, and supremum of affine functionals is a convex function.

Proof. Let $v_1, v_2 \in \Omega_K$ and $c \in [0, 1]$. Since $f(v_1) > 0$ and $f(v_2) > 0$ for all $f \in K$, and $c \geq 0$ and $1 - c \geq 0$, we conclude that

$$f(cv_1 + (1 - c)v_2) = cf(v_1) + (1 - c)f(v_2) > 0$$

for all $f \in K$. Hence $cv_1 + (1 - c)v_2 \in \Omega_K$. Similarly,

$$\begin{aligned} \Phi(cv_1 + (1 - c)v_2) &= \sup_{g \in K} g(cv_1 + (1 - c)v_2) \\ &= \sup_{g \in K} [cg(v_1) + (1 - c)g(v_2)] \\ &\leq c \sup_{g \in K} g(v_1) + (1 - c) \sup_{g \in K} g(v_2) \\ &= c\Phi(v_1) + (1 - c)\Phi(v_2), \end{aligned}$$

which proves the second part of the lemma. ■

10.1.3 Example: a convex function

Here is an example of how Lemma 10.1 can be used. Let us prove that the subset $\Omega = \mathbf{H}_+^n$ of the set $V = \mathbf{H}^n$ of Hermitian n -by- n matrices, consisting of all *strictly positive definite matrices*, is convex, and that the function $\Phi : \Omega \mapsto \mathbf{R}$, defined by

$$\Phi(M) = \text{trace}(M^{-1}),$$

is convex.

Note that doing this via the “positive eigenvalues” definition of positive definiteness would be difficult. Luckily, there is another definition: a matrix $M \in \mathbf{H}_+^n$ is positive definite if and only if $x'Mx > 0$ for all $x \in \mathbf{C}^n$, $x \neq 0$. Note that any $x \in \mathbf{C}^n$ defines an affine (actually, a linear) functional $f = f_x : \mathbf{H}^n \rightarrow \mathbf{R}$ according to

$$f_x(M) = x'Mx.$$

Hence, $\mathbf{H}_+^n$ is a subset of $\mathbf{H}^n$ defined by some (infinite) set of strict linear inequalities. According to Lemma 10.1, $\mathbf{H}_+^n$ is a convex set.

To show that Φ is a convex function, note that, for $M \in \mathbf{H}_+^n$, the identity

$$v'M^{-1}v = \sup_{u \in \mathbf{C}^n} \{2\text{Re}(v'u) - u'Mu\}$$

holds. Hence

$$\Phi(M) = \sum_{i=1}^n e_i'M^{-1}e_i = \sup_{u_i \in \mathbf{C}^n} \sum_{i=1}^n \{2\text{Re}(e_i'u_i) - u_i'Mu_i\}$$

is supremum of a set of affine functionals

$$g[u_1, \dots, u_n] : M \mapsto \sum_{i=1}^n \{2\text{Re}(e_i'u_i) - u_i'Mu_i\}.$$

According to Lemma 10.1, Φ is convex.

10.1.4 Second derivative and convexity

Let us call a function $f : \Omega \rightarrow \mathbf{R}$ defined on a subset Ω of $V = \mathbf{R}^n$ *twice differentiable* at a point $v_0 \in \Omega$ if there exists a symmetric matrix $W \in \mathbf{S}_{\mathbf{R}}^n$ and a row vector p such that

$$\frac{f(v) - f(v_0) - p(v - v_0) - 0.5(v - v_0)'W(v - v_0)}{\|v - v_0\|^2} \rightarrow 0 \text{ as } v \rightarrow v_0, v \in \Omega,$$

in which case $p = f'(v_0)$ is called the *first derivative* of f at v_0 and $W = f''(v_0)$ is called the *second derivative* of f at v_0 .

Lemma 10.2 *Let $\Omega \subset \mathbf{R}^n$ be a convex subset of $\mathbf{R}^n$. Let $f : \Omega \rightarrow \mathbf{R}$ be a function which is twice differentiable and has a positive semidefinite second derivative $W = f''(v_0) \geq 0$ at every point $v_0 \in \Omega$. Then f is convex, and can be defined as the maximum of affine functionals*

$$f(v) = \max_{u \in \Omega} \{f(u) + f'(u)(v - u)\}.$$

For example, let Ω be the positive quadrant in $\mathbf{R}^2$, i.e. the set of vectors $[x; y] \in \mathbf{R}^2$ with positive components $x > 0$, $y > 0$. Obviously Ω is convex. Let the function $f : \Omega \rightarrow \mathbf{R}$ be defined by $f(x, y) = 1/xy$. According to Lemma 10.2 is convex, because the second derivative

$$W(x, y) = \begin{bmatrix} d^2 f/dx^2 & d^2 f/dxdy \\ d^2 f/dydx & d^2 f/dy^2 \end{bmatrix} = \begin{bmatrix} 2/x^3y & 1/x^2y^2 \\ 1/x^2y^2 & 2/xy^3 \end{bmatrix}$$

is positive definite on Ω . Moreover, the identity

$$\frac{1}{xy} = \max_{x_1 > 0, y_1 > 0} \left\{ \frac{1}{x_1 y_1} - \frac{x - x_1}{x_1^2 y_1} - \frac{y - y_1}{x_1 y_1^2} \right\}$$

holds for $x, y > 0$.

10.1.5 Convexity-preserving operations

In addition to Lemma 10.2 and Lemma 10.1, which help establishing convexity “from scratch”, the following statements can be used to derive convexity of one function from convexity of other functions.

Lemma 10.3 *Let V be a vector space, $\Omega \subset V$.*

- (a) If $f : \Omega \rightarrow \mathbf{R}$ and $g : \Omega \rightarrow \mathbf{R}$ are convex functions then $h : \Omega \rightarrow \mathbf{R}$ defined by $h(v) = f(v) + g(v)$ is convex as well.
- (b) If $f : \Omega \rightarrow \mathbf{R}$ is a convex function and $c > 0$ is a positive real number then $h : \Omega \rightarrow \mathbf{R}$ defined by $h(v) = cf(v)$ is convex.
- (c) If $f : \Omega \rightarrow \mathbf{R}$ is a convex function, U is a vector space, and $L : U \rightarrow V$ is an affine function, i.e.

$$L(cu_1 + (1 - c)u_2) = cL(u_1) + (1 - c)L(u_2) \quad \forall c \in \mathbf{R}, u_1, u_2 \in U,$$

then the set

$$L^{-1}(\Omega) = \{u \in U : L(u) \in \Omega\}$$

is convex, and the function $f \circ L : L^{-1}(\Omega) \rightarrow \mathbf{R}$ defined by $(f \circ L)(u) = f(L(u))$ is convex.

For example, let $g : \mathbf{S}_{\mathbf{R}}^3 \rightarrow \mathbf{R}$ be defined on symmetric 2-by-2 matrices by

$$g\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = x^2 + y^2 + z^2.$$

To prove that g is convex, note that $g = f \circ L$ where $L : \mathbf{S}^3 \rightarrow \mathbf{R}^3$ is the affine (actually, linear) function defined by

$$L\left(\begin{bmatrix} x & y \\ y & z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

and $f : \mathbf{R}^3 \rightarrow \mathbf{R}$ is defined by

$$f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = x^2 + y^2 + z^2.$$

Lemma 10.2 can be used to establish convexity of f (the second derivative of f turns out to be the identity matrix). According to Lemma 10.3, g is convex as well.

10.1.6 The Hahn-Banach Theorem

The basis for all convex duality proofs is the fundamental *Hahn-Banach Theorem*. The theorem can be formulated in two forms: geometric (easier to understand) and functional (easier to prove).

By definition, an element v_0 of a real vector space V is called an *interior point* of a subset $\Omega \subset V$ if for every $v \in V$ there exists $\epsilon = \epsilon_v > 0$ such that $v_0 + tv \in \Omega$ for all $|t| < \epsilon_v$.

Theorem 10.1 *Let Ω be a convex subset of a real vector space V such that 0 is an interior point of Ω . If $v_0 \in V$ is not an interior point of Ω then there exists a linear function $L : V \mapsto \mathbf{R}$, $L \neq 0$, such that*

$$L(v_0) \geq \sup_{v \in \Omega} L(v).$$

In other words, a point not strictly inside a convex set can be separated from the convex set by a hyper-plane.

To give an alternative formulation of the Hahn-Banach Theorem, remember that a non-negative function $q : V \mapsto \mathbf{R}$ defined on a real vector space V is called a *semi-norm* if it is convex and positively homogeneous (i.e. $p(av) = ap(v)$ for all $a \geq 0$, $v \in V$).

Theorem 10.2 *Let $q : V \mapsto \mathbf{R}$ be a semi-norm on a real vector space V . Let V_0 be a linear subspace of V , and $h_0 : V_0 \mapsto \mathbf{R}$ be a linear function such that $q(v) \geq h_0(v)$ for all $v \in V_0$. Then there exists a linear function $h : V \mapsto \mathbf{R}$ such that $h(v) = h_0(v)$ for all $v \in V_0$, and $h(v) \leq q(v)$ for all $v \in V$.*

To relate the two formulations, define $q(v)$ as the *Minkovski' functional* of Ω :

$$q(v) = \inf\{t > 0 : t^{-1}v \in \Omega\},$$

and set

$$V_0 = \{tv_0 : t \in \mathbf{R}\}, \quad h_0(tv_0) = t.$$

10.1.7 Duality gap for linear programs

To demonstrate utility of the Hahn-Banach theorem, let us use it to prove the “zero duality gap” statement for linear programs.

Theorem 10.3 *Let A, B, C be real matrices of dimensions n -by- m , n -by-1, and 1-by- m respectively. Assume that there exists $v_0 \in \mathbf{R}^m$ such that $Av_0 < B$. Then*

$$\sup\{Cv : v \in \mathbf{R}^m, Av \leq B\} = \inf\{B'p : p \in \mathbf{R}^n, A'p = C', p \geq 0\}. \quad (10.1)$$

The inequalities $Av \leq B$, $Av_0 < B$, and $p \geq 0$ in (10.1) are understood component-wise. Note also that inf over an empty set equals *plus infinity*. This can be explained by the fact that inf is the *maximal lower bound* of a set. Since every number is a lower bound for an empty set, its infimum equals $+\infty$. Theorem 10.3 remains valid when there exist no $p \geq 0$ such that $A'p = C'$, in which case it claims that inequality $Av \leq B$ has infinitely many solutions, among which Cv can be made arbitrarily small.

Proof. The inequality

$$\sup\{Cv : v \in \mathbf{R}^m, Av \leq B\} \leq \inf\{B'p : p \in \mathbf{R}^n, A'p = C', p \geq 0\}$$

is straightforward: multiplying $Av \leq B$ by $p' \geq 0$ on the left yields $p'Av \leq B'p$; when $A'p = C'$, this yields $Cv \leq B'p$.

The proof of the inverse inequality

$$\sup\{Cv : v \in \mathbf{R}^m, Av \leq B\} \geq \inf\{B'p : p \in \mathbf{R}^n, A'p = C', p \geq 0\}$$

relies on the Hahn-Banach theorem.

Let y be an upper bound for Cv subject to $Av \leq B$. If $y = \infty$ then, according to the already proven inequality, there exist no $p \geq 0$ such that $A'p = C'$, and hence the desired equality holds.

If $y < \infty$, let e denote the n -by-1 vector with all entries equal to 1. Consider the set

$$\Omega = \left\{ x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} Cv - \delta + 1 \\ e - Av - \Delta \end{bmatrix} \in \mathbf{R}^{n+1} : \Delta > 0, \delta > 0 \right\}.$$

Then

- (a) Ω is a convex set (as a linear transformation image of a set defined by linear inequalities);
- (b) zero is an interior point of Ω (because it contains the open cube $|x_i| < 1$, which can be seen by setting $v = 0$);
- (c) vector $[y + 1; e - B]$ does not belong to Ω (otherwise $Av + \Delta = B$ and $Cv - \delta = y$, which contradicts the assumption that $Cv \leq y$ whenever $Av \leq B$).

According to the Hahn-Banach Theorem, this means that there exists a non-zero linear functional

$$L \begin{bmatrix} x_0 \\ \bar{x} \end{bmatrix} = L_0 x_0 + \bar{L}' \bar{x},$$

where $L_0 \in \mathbf{R}$, $\bar{L} \in \mathbf{R}^n$, defined on $\mathbf{R}^{n+1}$, such that

$$L_0(Cv - \delta + 1) + \bar{L}'(e - Av - \Delta) \leq L_0(y + 1) + \bar{L}'(e - B) \quad \forall \Delta > 0, \delta > 0, v. \quad (10.2)$$

Looking separately at the coefficients at v, δ, Δ and at the constant term in (10.2) implies

$$L_0 C = \bar{L}' A, \quad L_0 \geq 0, \quad \bar{L} \geq 0, \quad L_0 y \geq \bar{L}' B. \quad (10.3)$$

Note that L_0 cannot be equal to zero: otherwise $\bar{L}' A = 0$ and $\bar{L}' B \geq 0$, which, after multiplying $Av_0 < B$ by $\bar{L} \geq 0$, $\bar{L} \neq 0$ yields a contradiction:

$$0 = \bar{L}' Av_0 < \bar{L}' B \leq L_0 y = 0.$$

If $L_0 > 0$ then for

$$p = \bar{L}/L_0$$

conditions (10.3) imply

$$A'p = C', \quad p \geq 0, \quad B'p \leq y.$$

■

Index

- $RH^{q,m}$, 53
- $RH_-^{q,m}$, 53
- $L_{2e}(\mathbf{R}^m)$, 14
- $\mathbf{C}_r$, 21

- Algebraic riccati equation, *see* ARE
- almost all, 14
- ARE, 68, 72, *see* Riccati equation
 - stabilizing solution, 68

- change of coordinates, 31
- completion of squares
 - in optimal control, 66
- controllable
 - pair (A, B) , 32
 - state space model, 32
- controller
 - stabilizing, 37
- convex
 - function, 167
 - optimization, 167
 - quasi-convex function, 167
 - set, 167

- detectability, 34
- dissipation inequality, 75
- disturbance, 36

- entropy integral, 45
- essential limit, 14

- Fourier transform, 23
- Frobenius norm, 87
- function
 - real analytical, 22
- functional
 - affine, 168
 - linear, 168

- Gaussian distribution, 50
- Gaussian noise
 - continuous time, 51

- H-Infinity norm, 25
- H-Infinity optimization, 45
- H2 norm, 49
- H2 optimization, 51
 - abstract, 67, 79
- Hamiltonian matrix, 77
- Hamiltonian system, 80
- Hankel norm, 53
- Hankel optimal model reduction
 - setup, 53
- Hurwitz matrix, 34

- L2 norm, 49
- Laplace transform, 22
- largest singular number, 25
- linear time invariant, *see* LTI, *see* LTI
- LTI
 - linear time invariant, 8
- Lyapunov equation
 - continuous time, 50

- meromorphic, 23
- MIMO
 - multiple input, multiple output, 7
- minimal
 - state space model, 32
- model
 - state space, 30
- model reduction, 10

- observable
 - pair (C, A) , 32
 - state space model, 32
- operator norm, 25
- outer approximation, 18
- Parseval identity, 23
- plant, 36
- rational function
 - negative semidefinite, 73
- Riccati equation, 76
 - dual, 85
 - stabilizing solution, 76
- Riccati inequality
 - strict, 70
- sampling time, 15
- signal, 14
- signals
 - concatenation, 16
- singularity
 - control, 42
 - at $\omega = \infty$, 43
 - sensor, 42
- SISO, *see* single input single output
- spectral factorization, 73
- stabilizability, 34
- storage function, 75
- Storage functions, 69
- system
 - autonomous, 16
 - input/output, 16
 - L2 power gain, 17
 - stability, 17
- theorem
 - small gain, 20
- time
 - continuous, 14
 - transfer matrix, 24
 - order, 32
 - zero measure, 14
 - zero/pole cancellation, 40