Dissipative Systems and Stability

Peter Moylan

Newcastle, NSW, Australia

August 2014

peter@pmoylan.org
http://www.pmoylan.org

Preface

The theory of dissipative systems builds on a great deal of prior work, most notably on research on passive systems and on stability. It can be viewed as an extension of passive systems theory. It was J.C. Willems who introduced the label "dissipative" and provided the first and most basic stability results.

The material in this book comes largely from research done by the author in collaboration with D. Hill, mostly at the University of Newcastle, New South Wales, but also at the University of California, Berkeley. The results for large-scale interconnected systems, which in some ways form the central core of the book, were inspired by results by M. Vidyasagar on interconnected passive systems. Some of the later chapters contain previously unpublished material.

Newcastle, NSW, Australia *Peter Moylan*
August 2014

Contents

CHAPTER 1

Introduction

1. Background

A dissipative system is one that dissipates energy.

That, of course, is too imprecise a statement to base a proper theory on. Even so, it is a good starting point for thinking about what dissipativeness means.

Historically, the theory of dissipative systems comes from the theory of passive systems, which in turn owes much to electrical circuit theory. A passive circuit is one made from only passive components: resistors, inductors, and capacitors. To make the theory logically complete, we have to add transformers and gyrators to this list; these are two-port or multiport devices that are power-neutral, in that power out is always equal to power in. Resistors consume electrical energy and turn it into a non-electrical form, usually heat. They can never produce energy. (We exclude the case of resistors with negative resistance. These are not considered to be passive.) Inductors and capacitors can store energy, and later release it, but they can never release more energy than was supplied to them.

Negative-valued capacitors and inductors are not passive, because they can store negative energy. We shall later see that instead they have a property called cyclo-passivity.

It is a well-known result — it follows from Kirchhoff's laws — that any circuit made up of passive components is itself passive. For this result to make sense, however, it is necessary to define what we mean by a passive circuit, as distinct from a passive component. To do this, we model the circuit as something that has both internal components and external ports. Each port has a voltage and an incoming current. The product of port voltage and current is, of course, the power input at that port. If we sum this over all ports, we get the total power input to the circuit as a function of time. The integral of this power is the total energy input since the initial time. To avoid complications, we assume that there is no initial stored energy. Then we call the circuit passive iff

$$\int_0^T v(t)^T i(t)dt \geq 0$$

for all $T \geq 0$, where $v(t)$ is the vector of port voltages, $i(t)$ is the vector of corresponding port currents, and the superscript T indicates vector (or matrix) transpose.

In control and systems theory we use inputs u and outputs y, so that the passivity criterion becomes

$$\int_0^T y(t)^T u(t)dt \geq 0$$

This is consistent with the electrical circuit case if we define the port voltages to be the outputs and the port currents to be the inputs. (Or conversely; or even a mixed allocation.) The integral in question is still a measure of energy, and the passivity condition is still that the net energy transfer, from the initial time up to any arbitrary time T, be into the system (positive, or at least nonnegative). This does not rule out the integrand going negative some of the time. There can be an outward energy flow, as long as it does not exceed the energy previously put into the system.

It is, of course, possible to find other physical systems where the product of input and output has the dimensions of power. Consider, for example, a rotating machine where the input is the shaft torque and the output is the rotational speed. Such systems can be treated by exactly the same theory, without change. An example of the relevant theory is a result that says that a passive system, subject to certain assumptions, is stable. Much of this book will be about such results.

Now, it will turn out that none of that relevant theory will require that the "energy" in our calculations correspond to real physical energy. We can work with a product of input and output without any requirement that that product be anything more than a mathematical abstraction.

Having taken that step, we are in a position of being able to define an arbitrary *supply rate*: some function of input and output that can be more complicated than a simple product of input and output. That lets us say that a system is dissipative with respect to that supply rate if a certain "energy" inequality is satisfied. Naturally we do not require that this "energy" have anything to do with physical energy. We are happy to work at an abstract level provided that we can, at some stage, use these abstractions to deduce useful properties like stability.

The concept of a dissipative system is primarily due to J.C. Willems. The most readable approach to the Willems approach to dissipativeness can be found in [**Wil72**].

This book presents our approach to the theory of dissipative systems. We depart from the Willems approach in two ways:

(1) Willems defined dissipativeness in terms of the existence of an internal storage function, in addition to some other conditions. We prefer to define dissipativeness as a pure input-output property, and then to deduce the existence of a storage function as a consequence of that input-output property. This allows us to carry along a parallel development of both input-output and state-space properties.
(2) The bulk of our work concentrates on supply rates that are quadratic functions of input and output. There has been a certain amount of past research, including that of Willems, that has obtained useful results for linear systems and quadratic supply rates. What will be seen in this book is that even without linearity we can get useful results for the case of quadratic supply rates.

Despite these differences, much of the material in the early chapters overlaps the results of Willems. The differences become more important once we get to results about stability and instability.

Before proceeding, it is worth noting that the term "dissipative system" can also be found in the thermodynamics literature. This is a mere coincidence of terminology. The property discussed in connection with thermodynamics does not appear to have much of an overlap with the ideas of systems theory.

In the remainder of this chapter we will look at some of the ideas that led up to the notion of a dissipative system (in our sense). Chapter 2 will then pin down more precisely what we mean by dissipativeness.

2. The notion of a dynamical system

A *system*, in the formalism that we are going to use, is something that has an input u, an output y, and some sort of operator G that relates the output and input via the equation

$$y = Gu$$

The operator G can be linear or nonlinear. When it is nonlinear, it is more common to see the equation written as

$$y = G(u)$$

In this book, however, we will find it convenient to drop the parentheses except where absolutely needed to avoid confusion.

In the vast majority of applications of system theory, the signals u and y are functions of time. In that case, their values at time t are denoted by $u(t)$ and $y(t)$, respectively. Very occasionally, they are functions of something else. In picture processing, for example, our signals are functions over a two-dimensional space. In electrical field theory, the signals of interest are vector functions of both space and time. We should be able to allow any of these possibilities in a well-developed theory. That means that we have to introduce the concept of a *signal space*, which roughly speaking means the set of all possible signals for the system under consideration. We shall return to that detail in the next chapter.

Let us note, in passing, that it is not absolutely essential to divide up our signals into two classes called inputs and outputs. It would be entirely possible to develop a systems theory that simply called all external variables and possibly also the internal variables signals, without further discrimination. Most of the existing system theory does, however, make a distinction between inputs and outputs, so we will continue to look at a system as a map between inputs and outputs in this book. It should be noted, nevertheless, that in applications it often turns out that the decision as to whether a particular variable is an input or an output is an entirely arbitrary one.

The notation suggests that to each input there is a unique output. Actually, much of the theory works equally well if we define G to be a set of pairs (u, y), with no requirement of uniqueness. We shall not, however, take that step in this book, because the notation $y = Gu$ is so much more readable.

We call the operator G *memoryless* if the value of $y(t)$ depends only on the value of $u(t)$, and not on the past or future values of the input. Most of the time we shall not need to talk about memoryless operators, because nearly all interesting classes of systems have some memory. Even so, there are plenty of examples where, even though a system has memory, it still has some memoryless subsystems.

We call G *causal* if the present value of the output depends only on the past and present values of the input, and not on the future values. A more formal definition will be given in the next chapter. Most of the time we will be studying only causal systems.

3. The property of passivity

For the sake of having a concrete example, let us consider an electrical n-port circuit, where $u(t)$ is the vector of port voltages and $y(t)$ is the vector of port currents at time t. (This is not the only way to allocate inputs and outputs, but it is one common way.) With a suitable allocation of sign conventions, the power flowing into the circuit at time t is

$$p(t) = \sum u_i(t)y_i(t) = y(t)^T u(t)$$

where the superscript T represents vector transpose. Thus, the total energy into the circuit, over the time period t_0 to t_1, is

$$E(u, t_0, t_1) = \int_{t_0}^{t_1} y(t)^T u(t)dt$$

This formula is valid for any electrical n-port, whether linear or nonlinear, time-invariant or time-varying.

Traditionally a circuit is called passive if it contains no energy sources. It can dissipate energy by turning it into heat, and it can store energy in inductors and capacitors, but it cannot put out more energy than has been put into it. In other words, a passive circuit has the property that

$$E(u, t_0, t_1) = \int_{t_0}^{t_1} y(t)^T u(t) dt \geq 0$$

for any input u — that is, for any time history of input voltages — and for any t_0 and any $t_1 \geq t_0$. Although this inequality has been derived from the traditional definition of passivity, the modern tendency is to use the inequality as the definition of passivity.

For most of this book we will be concerned only with time-invariant systems, since those are the most common systems of interest. If a circuit or system is time-invariant then any choice of starting time is as good as any other. That means that the above condition can be simplified to

$$E(u, 0, T) = \int_{0}^{T} y(t)^T u(t) dt \geq 0$$

for any u and any $T \geq 0$. Of course, this works only in the time-invariant case.

It is tempting to simplify the condition even further to $E(u, 0, \infty) \geq 0$. That, however, would be an unjustified step. Although the infinite-time condition says that the circuit cannot ultimately deliver any net energy, it allows the possibility that initially some positive energy is produced. (But is later taken back.) We do not consider a circuit with that property to be passive. An example of such a circuit would be one built from positive resistors and negative capacitors. In a later chapter, we will call that sort of circuit *ultimately passive* or *cyclopassive*, but not passive.

4. Taking stored energy into account

Strictly speaking, the passivity condition of the last section should be called "external passivity". Alternatively, it should be qualified by a condition like "if there is no initial stored energy". If we make a circuit from passive components, but then charge up the capacitors before time 0, the circuit can of course deliver some energy after time 0.

In terms of definitions, there are two ways that one can treat that scenario. One way is to say that an initially energised circuit is not passive, even if it meets the passivity criteria in every other way. The other is to modify the defining inequality in such a way as to include an initial energy term. There are good arguments for both points of view. They are discussed in [**MCS82**], where a distinction is made between "a circuit" and "an instance of a circuit". The approach that will be taken in this book is to make a distinction between the concepts "internally passive" and "externally passive". We will, in fact, put considerable effort into looking at the connection between those two properties. When only an input-output model of the circuit or system is available — that is, when we know nothing about the internal state — then the external properties are the only things we can work with.

If, on the other hand, we have a state-space model, we can introduce the notion of a stored energy $\phi(x)$ that depends on the internal state x. Then, by an obvious extension of the arguments of the last section, the passivity definition becomes

$$\text{initial stored energy} + \text{energy input} \;\geq\; \text{final stored energy}$$

That is,

$$\phi(x(t_0) + \int_{t_0}^{t_1} y(t)^T u(t)dt \geq \phi(x(t_1)$$

This is the property that we will call internal passivity. Whether we work with internal or external passivity will depend on whether a state-space model is available.

In later chapters we will find it convenient to introduce a factor of 2 into the integrand in this inequality. (It will, it turns out, save us from having to put a factor of 1/2 into some of the other formulae.) The only thing that this will change will be to make the conceptual stored energy $\phi(x)$ twice the physical stored energy, a detail that is no more important than a decision to use a different set of units.

5. Passivity and stability

Informally, we call a system unstable if a bounded input can lead to an unbounded output. (More precise definitions will be saved for a later chapter.) Intuitively, we feel that this can happen only if the system is producing energy, or has some internal energy source; otherwise, the response is going to be damped. A more careful investigation shows that this intuitive judgement is very close to the truth. There are borderline cases that force us to require something called "strong passivity" rather than mere passivity, but with that proviso it is true that passivity implies stability.

The real strength of this result is the known fact that, if we connect a number of passive circuits together, the overall circuit is still passive. That potentially means that we can deduce the stability of a circuit that is too complicated for easy analysis. The strong passivity requirement adds some complications, but they are not major complications.

Once we move beyond electrical circuits, essentially all of the passivity theory continues to be valid. There are other physical systems where the product of input and output still represents power. (Think of the torque and speed of a rotating machine, for example.) More importantly, though, the theory that shows that passivity implies stability does not require that the integral in the defining inequality should have any meaning in terms of physical energy. The theory works perfectly well if the stored energy function $\phi(x)$ represents some sort of pseudo-energy that has nothing to do with real physical energy.

The idea that we can do something with pseudo-energy dates back to the second method of Lyapunov for proving stability. It is now very well understood that Lyapunov functions still work even if they have nothing to do with energy. All that is required is that they have the right mathematical properties.

Is an interconnection of passive systems itself passive? For electrical circuits, the answer is clear. For anything else, it depends on how we define "interconnection". In a later chapter we will look at what kinds of interconnection work, in terms of preserving the stability property.

6. Finite gain systems

Let us temporarily move away from passive systems, and look at an apparently unrelated property. For any given time $T \geq 0$, let u_T denote a truncated version of the input u, where $u(t)$ has been set to zero for all $t \geq T$; and similarly for y_T. Then we say that the system has finite gain if there exists a constant k such that

$$\|y_T\| \leq k\,\|u_T\|$$

for any u, and any $T \geq 0$. (Note that the same k is required to work for every T.) As usual the double lines denote a norm. This condition is often taken to be the definition of input-output stability, and indeed it is how we will define external stability in this book.

Strictly speaking, we cannot proceed without defining which norm we mean, because there are infinitely many ways to define a norm. In finite dimensional spaces this does not matter, because there is a sense in which all norms are equivalent to one another; but our inputs and outputs are functions of time, and therefore live in infinite-dimensional spaces. In practice, the answer is "the norm we used when defining the system model".

If finite gain is to be used as the definition of input-output stability, then of course finite gain implies input-output stability. That is not a very interesting result. There are other questions that are worth pursuing, though. Does finite gain stability imply internal stability? (The answer is yes, subject to some technical conditions.) Does an interconnection of finite gain systems still have the finite gain property? (No, in general, but we can find situations where it is preserved.)

7. The extension to dissipative systems

Let us digress slightly and consider the Nyquist stability criterion for a simple feedback loop. This criterion says, in essence, that the loop will become unstable when the loop gain becomes -1. (Which is effectively +1 when we consider that we are using negative feedback.) That is, when the magnitude of the loop gain becomes 1, and its phase becomes 180 degrees. If the loop gain never gets that far, the system is stable.

A result known as the small gain theorem, which also works for nonlinear systems, is based on ensuring stability by making sure that the magnitude of the loop gain is always less than 1.

At the other extreme, a linear passive system has the property that its phase shift is always between -90 and +90 degrees. Thus, two linear passive systems in tandem must have a phase shift between -180 and +180 degrees. The addition of a "strict passivity" condition can ensure that the phase shift lies strictly between these bounds. That means that the Nyquist condition is satisfied, even though we have looked only at the phase shift and ignored the magnitude.

As it happens, this result can be extended to nonlinear systems, even though we no longer have a good definition of phase shift. The result is then known as the "positive operator theorem".

An obvious question that now arises is whether we can find some sort of intermediate property where gain magnitude and phase can be traded off against each other. This would presumably produce a range of results, with passivity at one extreme of the spectrum, and finite gain at the other end.

One such result was introduced by Zames [**Zam66**] with his introduction of the notion of conicity. For linear systems, conicity is a property that confines a transfer function to avoid a circle in the complex plane. At one extreme the circle is centred at the origin and we have a finite gain condition. At the other extreme, the circle degenerates into a straight line, and we have a passivity condition.

Simultaneously and independently Sandberg [**San64**] came up with his own way of unifying the finite gain results and the passivity results. It is interesting to note that, although the final results in the Sandberg and Zames papers are roughly equivalent, the paths that they took to get those results were very different.

The results in the present book are in a sense extensions of the Zames and Sandberg results, but via yet another path. Our results can be seen as following on from work by J C Willems [**Wil72**], who defined the concept "dissipative" in a very general way.

Consider again the definition of finite gain. The signal norm that is probably the most used in systems theory is the L_2 norm. With this choice, the finite gain condition becomes

$$\int_0^T \left(-y(t)^T y(t) + k^2 u(t)^T u(t)\right) dt \geq 0$$

When we put it this way, the similarity to the passivity definition becomes a little clearer. We can include both possibilities if we allow an arbitrary quadratic function as the integrand. That is, we are motivated to look at the condition

$$\int_0^T \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t)\right) dt \geq 0$$

for some given matrices Q, S, and R, with Q and R symmetric. This, in fact, is the property that will be called (Q, S, R) dissipativeness in the remainder of this book.

Not all choices of Q, S, and R lead to interesting results. Consider, for example, the case where S is zero, and Q and R are positive definite. Here it is obvious that the inequality is satisfied regardless of what happens to the inputs and outputs. That means that it is not a property of the system being studied. What we need, instead, are the properties that

(1) If we could freely choose the values of u and y, then it would always be possible to find values of u and y such that the integrand goes negative.
(2) Nevertheless, the overall integral will still be nonnegative, because we can not freely choose u and y; our choices are constrained by the condition $y = Gu$.

If that is true, we can say that (Q, S, R) dissipativeness is a meaningful property, rather than merely being a property of the matrices. It should be clear that there is still plenty of room to find interesting families of Q, S, and R. For example, passivity is $(0, I, 0)$ dissipativeness, and finite gain is $(-I, 0, k^2 I)$ dissipativeness.

CHAPTER 2

Defining dissipativeness

1. Overview

The purpose of this chapter is to define what we mean by a dissipative system. To begin with, we need to look at some mathematical prerequisites.

2. The mathematical setting

The material in this section will already be familiar to most readers. Nevertheless, we need to define the scope of what we are talking about.

The signals — the inputs and outputs — that we need to work with will live in linear spaces. A linear space is a set of vectors, and two defined operations: addition of two vectors, and multiplication of a vector by a scalar. (We can skip the full formal definition, because it is available in a large variety of mathematical texts.) The scalars live in a field F, so we can describe our linear space as the pair (V, F), where V is the set of vectors and F is the field. In the majority of system theory applications the field is either the field of real numbers or the field of complex numbers, but we should not rule out other possibilities.

To work with concepts like stability, we need some notion of the size of a vector. The usual way of doing this is by introducing a norm. Let (V, F) be a linear space, where F is a field in which the inequality $|\alpha| \leq |\beta|$ can be given a meaning. (Not all fields have this property. Consider, for example, the field of integers modulo p, where p is a prime number.) Then we call the space a normed linear space if there exists a norm $\|\cdot\|$ mapping vectors to values in F, with the properties

(1) $\|v\| \geq 0$ for all v, and $\|v\| = 0$ iff $v = 0$.
(2) $\|\alpha v\| = |\alpha| \, \|v\|$ for any scalar α.
(3) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ (triangle inequality).

In fact we need something a little stronger than this: an inner product space. An inner product is a scalar-valued function of two vector variables, written $\langle \cdot, \cdot \rangle$, with the properties

(1) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
(2) $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$
(3) $\langle y, x \rangle = \langle x, y \rangle^*$, where the star denotes complex conjugate
(4) $\langle x, x \rangle \geq 0$ for all x, and $\langle x, x \rangle = 0$ iff $x = 0$.

Note that property 3 makes sense only if the underlying field is the field of real numbers or the field of complex numbers. Thus, we will not attempt to continue to allow more general fields.

It is easy to prove that the function

$$\|x\| = \sqrt{\langle x, x \rangle}$$

is a valid norm, usually called the induced norm. Therefore, every inner product space is a normed linear space.

For any linear operator Q, its adjoint Q^* is the linear operator defined by

$$\langle Q^* x, y \rangle = \langle x, Qy \rangle$$

In the case of matrices, the adjoint is just the complex conjugate of the transpose of the matrix. (Or, for real matrices, just the transpose.) A linear operator Q is called *self-adjoint* if $Q^* = Q$. There is no corresponding concept for nonlinear operators.

The notation $Q \geq 0$ will be used to mean $\langle y, Qy \rangle \geq 0$ for all y. We call such an operator nonnegative definite. (Some authors prefer the term "positive semidefinite".) If we have the stronger property $\langle y, Qy \rangle > 0$ for all $y \neq 0$ then we call Q positive definite. The properties "negative definite" and "nonpositive definite" are defined similarly.

For a finite-dimensional Euclidean space, the obvious choice of inner product is the well-known dot product of two vectors, commonly written in matrix notation as $x^T y$, or $x^* y$ if we are working with complex numbers. (This time, the superscript star denotes the complex conjugate transpose of a matrix or vector.) In most of our applications, though, the vector spaces of interest are function spaces. A very common choice of function space is the space $L_2[0, \infty)$, where the inner product is defined as

$$\langle f, g \rangle = \int_0^\infty f(t)^T g(t) dt$$

That is if we are working with continuous-time systems. For discrete-time systems, a suitable choice of inner product is

$$\langle f, g \rangle = \sum_{k=0}^{\infty} f(k)^T g(k)$$

That serves as a reminder that we also need to define a *time line*, being the independent variable for the functions in our function space. This is usually the continuous half interval $[0, \infty)$ or its discrete-time equivalent. There are, however, some applications where it makes more sense to define the time line as $(-\infty, \infty)$.

Having defined a time line, we need to define a thing called the causal truncation operator P_T. This is, in the most general case, an orthogonal projection operator — that is, a linear operator with the properties $P_T^* = P_T$ and $P_T^2 = P_T$ — from the signal space to itself. Since we are going to use it to define causality, a useful choice of projection is

$$(P_T f)(t) = \begin{cases} f(t), \text{ for } t < T \\ 0, \text{ for } t \geq T \end{cases}$$

Note that this works equally well in continuous time or discrete time.

We are going to restrict our attention to causal systems. A system is called *causal* if the output at any time is independent of future inputs. Equivalently, it is causal iff the output up until time T is the same for two inputs that are identical until time T. That is, if

$$P_T G u = P_T G P_T u \text{ for all } u$$

We usually write this as

$$P_T G = P_T G P_T$$

We call the system *anticausal* if it would be causal with the time direction reversed. That is, if the output depends only on the future input. This is not a very interesting property in itself, since it is rare to meet an anticausal system, but it paves the way for our next definition: G is *memoryless* iff it is both causal and anticausal.

It is actually more common to define causality by the condition

$$\bar{P}_T G = \bar{P}_T G \bar{P}_T$$

where

$$\left(\bar{P}_T f\right)(t) = \begin{cases} f(t), \text{ for } t \leq T \\ 0, \text{ for } t > T \end{cases}$$

The difference, of course, lies in how we treat the signals at time $t = T$. These two definitions of causality are in fact equivalent, although it takes some thought to see why. For continuous-time systems, the difference between P_T and $\bar{P}_T$ is of only minor significance. There is a bigger difference in the case of discrete-time system models. Our preference for defining P_T the way we do is that it simplifies the notation in the discrete-time case.

The frequent requirement for dealing with truncated signals makes it convenient to define a truncated inner product

$$\langle x, y \rangle_T = \langle P_T x, P_T y \rangle$$

The fact that P_T is both self-adjoint and idempotent means that

$$\langle P_T x, P_T y \rangle = \langle P_T x, y \rangle = \langle x, P_T y \rangle$$

That is, we can truncate either or both of x and y, and get the same result.

3. Signal spaces

Since we are going to use inner products extensively in what follows, it is tempting to define a (linear or nonlinear) system G as a map from an inner product space U to another inner product space Y. It turns out, however, that this is not good enough. For one thing, it excludes the possibility of persistent inputs such as sine waves, because they would have infinite norms for most choice of norm. Secondly, it is hard to discuss stability in such a setting. Stability means, roughly speaking, that bounded inputs should produce bounded outputs. (There are several good alternative ways of making this more precise.) That means that we have to be able to consider the possibility that, for an unstable system, an input in U could produce an output that is not in Y.

The way to resolve this is to define a larger space

$$U_e = \{u : P_T u \in U \text{ for all } T\}$$

and similarly for the output space Y_e. The subscript "e" stands for "extended". Now our signals do not have to have finite norm, as long their causal truncation has finite norm. Obviously the norm of the truncated signal will grow with T, possibly without bound, but that is acceptable.

Now we can say that G maps U_e to Y_e. The smaller spaces U and Y are often called the spaces of small signals.

Stability is now easy to define. We say that G is input-output stable if $u \in U$ implies $Gu \in Y$, and input-output unstable if there is at least one input that violates this condition. It is convenient to define the set

$$K(G) \triangleq \{u \in U : Gu \in Y\}$$

With this notation, G is input-output stable iff $K(G) = U$.

Let us note that in this book we will be considering only time-invariant systems. That is, a time shift in the input will have no effect on the output except to shift it by the same amount. It is not difficult to extend the theory to time-varying systems, but the notation becomes more complicated, and some of the theorems require some side conditions to rule out awkward behaviour.

4. The definition of dissipativeness

Finally, we are in a position to define what we mean by saying that a system is (Q, S, R) dissipative.

Definition 1. *Let Q, S, and R be memoryless linear operators, with Q and R self-adjoint. Then the system defined by the (linear or nonlinear) operator equation $y = Gu$ is (Q, S, R) dissipative iff*

$$\langle Gu, QGu \rangle_T + 2 \langle Gu, Su \rangle_T + \langle u, Ru \rangle_T \geq 0$$

for all $u \in U_e$ and for all T.

It is usually more convenient to write the inequality as

$$\langle y, Qy \rangle_T + 2 \langle y, Su \rangle_T + \langle u, Ru \rangle_T \geq 0$$

where the constraint $y = Gu$ is implied. Note, however, that it would be perfectly possible to violate the inequality if we were able to choose arbitrary u and y which were not related by $y = Gu$.

Note that the definition is rather more restrictive than than of Willems [**Wil72**], who defined dissipativeness via a more general function of input and output. We are choosing to sacrifice some of the generality for the sake of getting more explicit results. By making the quantity on the left of the inequality a quadratic function of input and output we get a property that is, as will be seen in later chapters, easier to test. There will be other benefits, such as stability tests that reduce down to simple matrix algebra.

5. Ultimate virtual dissipativeness and weak dissipativeness

At times it will be useful to refer to a couple of weaker dissipativeness properties.

Definition 2. *Let Q, S, and R be memoryless linear operators, with Q and R self-adjoint. Then the system defined by the (linear or nonlinear) operator equation $y = Gu$ is (Q, S, R) ultimately virtually dissipative (UVD) iff*

$$\langle Gu, QGu \rangle + 2 \langle Gu, Su \rangle + \langle u, Ru \rangle \geq 0$$

for all $u \in K(G)$.

The thing that makes this a weaker property is that we only require the inequality to hold for small signals, and there is no condition placed on the truncated-signal behaviour. This property will later be used to prove some instability results.

Another motivation for introducing the UVD property is that it is often convenient to work out whether a system is dissipative via a two-step process: first check the UVD condition, and then check for a side condition that will strengthen the property to full dissipativeness.

Another weak property will occasionally be useful.

Definition 3. *Let Q, S, and R be memoryless linear operators, with Q and R self-adjoint. Then the system defined by the (linear or nonlinear) operator equation $y = Gu$ is (Q, S, R) weakly dissipative iff there exists a constant β such that*

$$\langle Gu, QGu \rangle_T + 2 \langle Gu, Su \rangle_T + \langle u, Ru \rangle_T + \beta \geq 0$$

for all $u \in U_e$ and for all T.

This property will be useful for proving a weaker form of stability.

Observe that both of these properties are weaker than dissipativeness. A dissipative system is both UVD and weakly dissipative, but the converse is not true.

We have not yet introduced the notion of an internal state — that will be covered in the next chapter — but a brief digression into the role of an initial state might help explain the motivation for the weak form of dissipativeness. So far we have, as is standard in input-output descriptions, implicitly assumed that the equation $y = Gu$ is a description of the input-output behaviour when the system is started in some sort of standard rest condition. This is the best we can do if we know nothing about the internal state. If we do know about an initial state x_0, and we have the ability to set a value for x_0 before applying an input, then we should really be talking about a whole family, parametrised by x_0, of input-output behaviours $y = G(x_0)u$. For the purpose of the definitions in this chapter, each $G(x_0)$ is a different system.

It is conceivable, then, that $G(x_0)$ is (Q, S, R) dissipative but $G(x_1)$ is not, for any $x_1 \neq x_0$. Luckily the situation is a little simpler than that. It will turn out that if $G(x_0)$ is (Q, S, R) dissipative, then $G(x_1)$ is (Q, S, R) weakly dissipative for all states x_1 that are reachable from x_0. It will also turn out that the β in the definition of weak dissipativeness is a function of the initial state. Of course, to reach these conclusions we must have a system model that includes an initial state, and we must have a property called reachability. This will be explored further in the next chapter. For now it suffices to note that weak dissipativeness is very closely related to dissipativeness.

Ultimate virtual dissipativeness is quite a different matter. If a system is (Q, S, R) ultimately virtually dissipative but not (Q, S, R) dissipative, and we have access to an internal state space model, then that system will have some states that have negative stored energy.

CHAPTER 3

Storage Functions

1. Dissipativeness as a state-space property

The property of dissipativeness, as presented in the previous chapter, is an input-output property. Some people would prefer it to be a state-space property. The focus of this chapter is on showing that it can be both. Under suitable assumptions, external dissipativeness implies internal dissipativeness.

2. The state-space model

With an input-output model, we assume an operator that, in effect, produces an output at each time that is a function of the past and present history of the input. The state-space approach is to have a state $x(t)$ at time t that encodes what needs to be known about the past history; and then the output at time t can be written as a function of the present value of the state and the present value of the input.

The usual way of setting this up is to assume a metric space X (the state space), and a transition map $\psi(t_0, t_1, x_0, u)$ that gives the new state at time t_1 as a function of an initial state $x(t_0) = x_0$ and of the input history u over the time interval from t_0 to t_1. We require the transition map to satisfy the following properties.

- The limit $x(t) = \lim_{t_0 \to -\infty} \psi(t_0, t, 0, u)$ is in X for all t and all u. (We then call $x(t)$ the state at time t.)
- (Causality) $\psi(t_0, t_1, x_0, u_1) = \psi(t_0, t_1, x_0, u_2)$ for all $t_1 \geq t_0$, all $x \in X$, and all u_1, u_2 such that $u_1(t) = u_2(t)$ in the interval $t_0 \leq t \leq t_1$.
- (Initial state consistency) $\psi(t_0, t_0, x_0, u) = x_0$ for all values of these variables.
- (Semigroup property) $\psi(t_1, t_2, \psi(t_0, t_1, x_0, u), u) = \psi(t_0, t_2, x_0, u)$.
- (Unbiasedness) $\psi(t_0, t, 0, 0) = 0$ for all $t \geq t_0$.
- (Time invariance) $\psi(t_0 + T, t_1 + T, x_0, u_1) = \psi(t_0, t_1, x_0, u_2)$ for all $t_1 \geq t_0$, for all T, and all u_1, u_2 such that $u_2(t) = u_1(t + T)$.

We also need a readout map: a function r that lets us write the output as $y(t) = r(x(t), u(t))$. Naturally, we require this end result to be consistent with the input-output relation. Note that $y(t)$ depends only on the present value of state and input, and not their past history. It is the job of the state to capture what needs to be known about the past history.

The first property mentioned above implicitly assumes a time line that begins in the infinite past, with zero initial state at the infinite past. Another reasonable approach would be to define the time line to be only semi-infinite, starting at $t = 0$, but then we have to face the question of how the input-output relation is affected by the initial state. There are two good ways to look at that issue:

(1) We could assume that the input-output relation specified by $y = Gu$ describes only the response for zero initial state. That is, the input-output description is only for a "system initially at rest" condition. The state can be made to be nonzero only by applying an input.

(2) We could allow arbitrary initial states, but change the input-output equation to a more general map $y = G(x_0)u$.

This second possibility is the one usually preferred. In that case, however, it usually turns out that, if $G(0)$ is dissipative, $G(x_0)$ is only weakly dissipative for $x_0 \neq 0$. The reason for that will become clear as we proceed through this chapter. Dissipativeness in the input-output sense means that the system cannot produce energy from nowhere; the existence of initial stored energy destroys that property.

In what follows, we shall occasionally need to refer to reachable and/or controllable states. These are defined as follows.

Definition 4. *A state x_0 is controllable at time t_0 if there exists a $t_1 \geq t_0$ and a u such that $\psi(t_0, t_1, x_0, u) = 0$.*

Definition 5. *A state x_0 is reachable at time t_0 if there exists a $t_{-1} \leq t_0$ and a u such that $\psi(t_{-1}, t_0, 0, u) = x_0$.*

For a time-invariant system, the condition "at time t_0" is of course redundant in these definitions. Controllability refers to the ability to force the state to the origin from a given state. Reachability refers to the ability to get from the origin to a given state. For linear time-invariant finite-dimensional continuous-time systems these two concepts can be shown to be equivalent. More generally, they need not be equivalent.

For the definition of dissipativeness in a state-space context, we will choose a square-integral or sum-of-squares inner product. That means that we will be working with an "energy input" function

$$E(x_0, u, y, t_0, t_1) = \int_{t=t_0}^{t_1} \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t) \right) dt$$

for the continuous-time case, and

$$E(x_0, u, y, t_0, t_1) = \sum_{t=t_0}^{t_1 - 1} \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t) \right)$$

for the discrete-time case. The initial state $x(t_0) = x_0$ needs to be included in the notation because the relationship between u and y depends on the initial state.

With this definition, the (external) dissipativeness condition becomes

$$E(0, u, y, 0, T) \geq 0 \text{ for all } u \text{ and all } T \geq 0$$

Observe that, for the discrete-time case, we do not include the time t_1 in the sum. There is a reason for this. The main results in this chapter require the energy function to have a time separation property

$$E(x_0, u, y, t_0, t_2) = E(x_0, u, y, t_0, t_1) + E(x\left(t_1\right), u, y, t_1, t_2)$$

where $t_0 \leq t_1 \leq t_2$. If the sum included the values at both boundary times, we would be double-counting the value at time t_1, so that the time separation property would not hold. This, by the way, is the reason for defining the causal truncation operator P_T the way that was given in Chapter 2.

In continuous time the double-counting issue does not arise except when there is a delta function, or something similar, at time t_1. Even so, in case of doubt we have to remember that the integral notation must be read as "up to but not including the final time".

For the material of the following sections we need one extra definition.

Definition 6. *The system G(0) is* cyclodissipative *iff*

$$E(0, u, y, 0, T) \geq 0$$

for all u and all $T \geq 0$ such that $x(T) = 0$.

This is a weaker property than dissipativeness, because the inequality is required to hold only for that subset of inputs that return the state back to the origin. It will later be shown that, if this condition holds for zero initial and final state, then it holds for any arbitrary initial state, provided that the final state is equal to the initial state. The "cyclo" part of the name comes from the fact that we are requiring a certain behaviour for those inputs that return the state back to its initial value.

3. Storage functions

Note that the dissipativeness condition assumes a zero initial state. To allow for arbitrary initial states, we need the concept of internal stored energy.

DEFINITION 7. *A function $\phi : X \to R$ is called a* virtual storage function *for the system G iff $\phi(0) = 0$ and*

$$\phi\left(x\left(t_0\right)\right) + E\left(x\left(t_0\right), u, y, t_0, t_1\right) \geq \phi\left(x\left(t_1\right)\right)$$

for all $t_1 \geq t_0$, all $x\left(t_0\right)$, and all u, where y is the output that results from initial state $x\left(t_0\right)$ and input u. It is called a storage function *if in addition $\phi(x) \geq 0$ for all x.*

Note that this definition applies to any system, not just dissipative systems. We have not, however, specified that a (virtual) storage function must exist; and indeed, it turns out that in the general case no such function need exist. For (cyclo)dissipative systems, we can prove that such functions exist by exhibiting some suitable candidates.

DEFINITION 8. *The* required supply *is defined by*

$$\phi_r(x_0) = \inf_{u \in U_e, t_{-1} \leq t_0} E(0, u, y, t_{-1}, t_0)$$

with boundary conditions $x(t_{-1}) = 0$, $x(t_0) = x_0$.

DEFINITION 9. *The* virtual available storage *is defined by*

$$\phi_a^*(x_0) = -\inf_{u \in U_e, t_1 \geq t_0} E(x_0, u, y, t_0, t_1)$$

with boundary conditions $x(t_0) = x_0$, $x(t_1) = 0$.

There is no guarantee that either of these functions has finite values for any x. Let us therefore assign the value $+\infty$ to the "inf" if the boundary conditions cannot be met; and the value $-\infty$ if the boundary conditions can be met, but there is no lower bound on the value.

The question of interest is therefore under what conditions these functions have finite values. Let us begin with an obvious result.

LEMMA 1. *Regardless of dissipativeness or cyclodissipativeness,*

(1) *$\phi_r(x_0) < \infty$ for all reachable x_0; and*
(2) *$\phi_a^*(x_0) > -\infty$ for all controllable x_0.*

PROOF. Directly from the definitions of controllable and reachable states. □

We get a more useful result for a cyclodissipative system.

LEMMA 2. *If the system is cyclodissipative, then*

$$\phi_a^*(0) = \phi_r(0) = 0$$

and

$$\phi_r(x) \geq \phi_a^*(x) \text{ for all } x$$

PROOF. Suppose first that x_0 is both controllable and reachable. That means that we can find times $t_{-1} \leq t_0 \leq t_1$ such that we can construct at least one trajectory that passes through $x(t_{-1}) = 0$, $x(t_0) = x_0$, and $x(t_1) = 0$. For any such trajectory, cyclodissipativeness implies

$$E(0, u, y, t_{-1}, t_0) + E(x_0, u, y, t_0, t_1) \geq 0$$

Taking the infimum of both terms, we conclude that

$$\phi_r(x_0) - \phi_a^*(x_0) \geq 0 \text{ for any } x_0$$

For the special case $x_0 = 0$, it is clear that we can take the state from 0 to 0 with zero cost, so from the definition of ϕ_r it follows that $\phi_r(0) \leq 0$. By the same reasoning, we conclude that $\phi_a^*(0) \geq 0$. Combining these inequalities, the only possibility is $\phi_a^*(0) = \phi_r(0) = 0$.

If x_0 is uncontrollable or unreachable, the result still holds in a formal sense, since then we have either $\infty = \phi_r(x_0) \geq \phi_a^*(x_0)$ or $\phi_r(x_0) \geq \phi_a^*(x_0) = -\infty$. □

Combining the results of the last two lemmas, we conclude that a cyclodissipative system has the property

$$-\infty < \phi_a^*(x_0) \leq \phi_r(x_0) < \infty$$

for all x_0 that are both controllable and reachable.

Now, let us consider a third candidate for the name stored energy.

DEFINITION 10. *The* available storage *is defined as*

$$\phi_a(x_0) = -\inf_{u \in U_e, t_1 \geq t_0} E(x_0, u, y, t_0, t_1)$$

with boundary conditions $x(t_0) = x_0$, $x(t_1)$ unconstrained.

The only difference between the definitions of ϕ_a and ϕ_a^* is in the final boundary condition. This observation leads immediately to the conclusions that $\phi_a(x) \geq 0$ and $\phi_a(x) \geq \phi_a^*(x)$ for all x for which the functions are defined. These properties are independent of cyclodissipativeness. Note, however, that we have no finite upper bound for the available storage, even if the system is cyclodissipative. To get a upper bound, we need the system to be dissipative.

LEMMA 3. *If the system is dissipative, then*

$$\phi_a(0) = \phi_r(0) = 0$$

and

$$\phi_r(x) \geq \phi_a(x) \geq 0 \textit{ for all } x$$

PROOF. The proof is similar to the proof of the previous lemma. If x_0 is reachable, then we can find times $t_{-1} \leq t_0$ and an input u that takes the state from $x(t_{-1}) = 0$ to $x(t_0) = x_0$, and then to an unspecified $x(t_1)$. Dissipativeness then implies that

$$E(0, u, y, t_{-1}, t_0) + E(x_0, u, y, t_0, t_1) \geq 0$$

for any $t_1 \geq t_0$, regardless of $x(t_1)$. (It is the fact that this inequality holds for any $x(t_1)$ that makes the dissipative case different from the merely cyclodissipative case.) Taking the infimum of both terms, we conclude that

$$\phi_r(x_0) - \phi_a(x_0) \geq 0 \text{ for any } x_0$$

The fact that t_1 can be freely chosen means that

$$-\phi_a(x_0) = \inf_{u \in U_e, t_1 \geq t_0} E(x_0, u, y, t_0, t_1) \leq E(x_0, u, y, t_0, t_0) = 0$$

from which it follows that $\phi_a(x_0) \geq 0$. Note that this part of the argument is valid whether or not x_0 is reachable.

As for the cyclodissipative case, we can then argue that $\phi_r(0) \leq 0$ and then that $\phi_a(0)$ and $\phi_r(0)$ must both be 0.

That leaves only the case where x_0 is unreachable. In that case, of course, we have $\infty = \phi_r(x_0) \geq \phi_a(x_0) \geq 0$, which completes the proof. □

So far we have shown that the three potential storage functions have some interesting properties, but we have not yet shown that they are (virtual) storage functions. The next result provides the essential connection.

LEMMA 4. *The functions ϕ_a^* and ϕ_r are virtual storage functions for a cyclodissipative system. For a dissipative system, the functions ϕ_a and ϕ_r are storage functions.*

PROOF. We shall show the method of proof for ϕ_r only; the other cases are similar. First, recall that it has already been shown that $\phi_r(0) = 0$ in both the cyclodissipative and the dissipative cases. Next, consider two reachable states x_0 and x_1, and three times $t_{-1} \leq t_0 \leq t_1$. From the definition of ϕ_r, we have

$$\phi_r(x_1) \leq E(0, u, y, t_{-1}, t_1)$$

for any u that takes the state from $x(t_{-1}) = 0$ to $x(t_1) = x_1$, including those that takes the state via $x(t_0) = x_0$. For those u, we can break the energy input into two time intervals, giving

$$\phi_r(x_1) \leq E(0, u, y, t_{-1}, t_0) + E(x_0, u, y, t_0, t_1)$$

which may be rearranged as

$$E(0, u, y, t_{-1}, t_0) \geq -E(x_0, u, y, t_0, t_1) + \phi_r(x_1)$$

That means that

$$\phi_r(x_0) = \inf_{u \in U_e, t_{-1} \leq t_0} E(0, u, y, t_{-1}, t_0) \geq -E(x_0, u, y, t_0, t_1) + \phi_r(x_1)$$

from which it follows that

$$\phi_r(x_0) + E(x_0, u, y, t_0, t_1) \geq \phi_r(x_1)$$

This is our desired result, at least for reachable states. If x_0 and/or x_1 is unreachable the inequality is still true, but is less meaningful because one or more of the terms will be infinite. □

Note that the result for ϕ_r holds in both the cyclodissipative and dissipative cases. The only difference is that in the dissipative case we also have $\phi_r(x_0) \geq 0$. In the cyclodissipative case there is no sign constraint on the virtual storage function.

4. The relationship between external and internal dissipativeness

The results of the previous section lead immediately to the two key theorems of this chapter.

THEOREM 1. *A system is dissipative iff there exists a storage function ϕ, with $0 \leq \phi(x) < \infty$ for all reachable x.*

PROOF. If the system is dissipative, then the previous lemmas show that both $\phi_a(x)$ and $\phi_r(x)$ are storage functions satisfying the required conditions.

For the converse, suppose that some storage function ϕ exists. Setting $x_0 = 0$ in the definition of a storage function, we have

$$E(x(t_0), u, y, t_0, t_1) \geq \phi(x(t_1))$$

and then the result follows from the fact that $\phi(x(t_1)) \geq 0$. □

The point of this theorem is that it shows the connection between the input-output property of (external) dissipativeness and the state-space property of having a storage function. It is worth recalling here a point that has been suppressed in the notation. If $G(x_0)$ is the nonlinear operator representing the input-output map for this system, then what we have actually shown is that $G(0)$ is dissipative iff there exists a storage function for the state-space representation. If we start from a non-zero initial state — and therefore a nonzero initial stored energy — then the corresponding inequality is

$$E\left(x\left(t_0\right), u, y, t_0, t_1\right) \geq \phi\left(x\left(t_1\right)\right) - \phi\left(x\left(t_0\right)\right) \geq -\phi\left(x\left(t_0\right)\right)$$

from which we conclude that $G(x_0)$ is weakly dissipative. That clarifies a related point. In the definition of weak dissipativeness

$$E\left(x_0, u, y, t_0, t_1\right) + \beta \geq 0$$

the transition to a state-space model shows us that the "constant" β is a function of the initial state x_0. Putting this another way: if $G(0)$ is dissipative then $G(x_0)$ is weakly dissipative for all reachable x_0, but there is no single β that will work for all cases; it is a different form of weak dissipativeness for each separate initial state.

The corresponding result for cyclodissipative systems should by now be obvious.

Theorem 2. *A system is cyclodissipative iff there exists a virtual storage function ϕ, with $-\infty < \phi(x) < \infty$ for all $x \in X$ which are both controllable and reachable.*

Proof. If the system is cyclodissipative, then the previous lemmas show that both $\phi_a^*(x)$ and $\phi_r(x)$ are virtual storage functions satisfying the required conditions.

For the converse, suppose that some virtual storage function ϕ exists. From the definition of a virtual storage function, we known that $\phi(0) = 0$. Setting $x(t_0) = x(t_1) = 0$ in the definition of a storage function, we have

$$0 + E\left(0, u, y, t_0, t_1\right) \geq 0$$

for all trajectories that take us back to the origin. □

This leads to an interesting corollary.

Theorem 3. *If $G(0)$ is cyclodissipative, then $G(x_0)$ is also cyclodissipative, for any x_0 which is both controllable and reachable.*

Proof. If $G(0)$ is cyclodissipative, then there exists a virtual storage function ϕ which takes on finite values for states which are both controllable and reachable. Then, for any trajectory that takes us from initial state x_0 to the same final state, we have

$$\phi\left(x_0\right) + E\left(x_0, u, y, t_0, t_1\right) \geq \phi\left(x_0\right)$$

from which the result is obvious. □

This means that, unlike the dissipative case, there is no need to define a "weak" counterpart to the cyclodissipativeness property. The energy inequality applies for *any* cyclic motion that returns to its initial state.

Note the distinction between ϕ_a^* and ϕ_a. Since every dissipative system is also cyclodissipative, we can assert that ϕ_a^* exists for a dissipative system, although we cannot conclude anything about its sign properties. For an arbitrary cyclodissipative system, however, we cannot even be sure that ϕ_a exists. Actually, we can be certain that ϕ_a does not exist for a system that is cyclodissipative but not dissipative, since its existence would lead to the conclusion that the system was dissipative.

There is still one gap in our results. For dissipative systems, there is a clear connection between external properties and state-space properties. Cyclodissipativeness is, however, defined in such a way that requires a state-space model; it is not obvious how one could define it in a way that did not mention states. The connection, it turns out, requires a weak observability property.

THEOREM 4. *Suppose that $G(0)$ is unbiased ($G(0)0 = 0$) and that $G(0)$ has a state-space realisation that is observable in the sense that $u \in K(G(0))$ implies that the zero-initial-state response satisfies $\lim_{t\to\infty} x(t) = 0$. Then ultimate virtual dissipativeness (UVD) of $G(0)$ is equivalent to cyclodissipativeness of that state-space realisation.*

PROOF. Suppose that $G(0)$ is UVD. The UVD definition can be written as

$$E\left(0, u, y, t_0, \infty\right) \geq 0$$

for all $u \in K(G(0))$. Now, consider any control u that takes the state from $x(t_0) = 0$ to $x(t_1) = 0$, with $u(t) = 0$ for all $t > t_1$. Unbiasedness means that $x(t)$ and $y(t)$ also remain zero for $t > t_1$. That means that

$$E\left(0, u, y, t_0, t_1\right) \geq 0$$

Since u is arbitrary in the time interval $[t_0, t_1]$, apart from the boundary conditions on the state, this implies that the system is cyclodissipative.

For the converse, cyclodissipativeness implies

$$E\left(0, u, y, t_0, t_1\right) \geq 0$$

for any $u \in K(G(0))$ such that $x(t_1) = 0$. Letting $t \to \infty$, we get the UVD condition. □

5. Bounds on storage functions

So far we know that a dissipative system has at least one storage function, and that a cyclodissipative system has at least one virtual storage function. When we consider how to calculate the storage functions — a topic that will be covered in a later chapter — we will find that there are, in general, multiple solutions. In fact, it turns out that if there is more than one solution then there is an infinity of solutions.

Let us begin with an elementary result.

THEOREM 5. *If ϕ_1 and ϕ_2 are storage functions for a dissipative system (virtual storage functions for a cyclodissipative system), then $\alpha\phi_1 + (1-\alpha)\phi_2$ is also a (virtual) storage function, for any $\alpha \in [0..1]$.*

PROOF. From the definition of (virtual) storage functions, we have

$$\alpha\phi_1\left(x\left(t_0\right)\right) + \alpha E\left(x\left(t_0\right), u, y, t_0, t_1\right) \geq \alpha\phi_1\left(x\left(t_1\right)\right)$$

and

$$(1-\alpha)\phi_2\left(x\left(t_0\right)\right) + (1-\alpha)E\left(x\left(t_0\right), u, y, t_0, t_1\right) \geq (1-\alpha)\phi_2\left(x\left(t_1\right)\right)$$

The result then follows by adding these two inequalities. □

This tells us that the set of all storage functions is a convex set, and the set of all virtual storage functions is a convex set. Furthermore, since a dissipative system is also cyclodissipative, the set of all storage functions for a dissipative system is a convex subset of the convex set of all of its virtual storage functions.

We do not have, as yet, a good characterisation of the boundaries of this set. We can, however, know the maximum and minimum elements of the set. This is shown by the following theorem.

THEOREM 6. *If ϕ is a storage function for a dissipative system, then*

$$0 \leq \phi_a(x) \leq \phi(x) \leq \phi_r(x)$$

for all $x \in X$.

PROOF. For the upper bound, let x_1 be any reachable state. (For unreachable states, there is of course nothing to prove.) Let u be any control taking $x(t_0) = 0$ to $x(t_1) = x_1$. Since $\phi(x(t_0)) = 0$, we have

$$E(0, u, y, t_0, t_1) \geq \phi(x_1)$$

Since this is true for any u meeting the boundary conditions, and any t_1, we can conclude that

$$\phi(x_1) \leq \inf_{u \in U_e} E(0, u, y, t_0, t_1) = \phi_r(x_1)$$

For the lower bound, consider instead those trajectories starting from $x(t_0) = x_0$. This time we have

$$\phi(x_0) + E(x_0, u, y, t_0, t_1) \geq \phi(x_1) \geq 0$$

so that

$$\phi(x_0) \geq -E(x_0, u, y, t_0, t_1)$$

The conclusion then follows from the definition of ϕ_a. □

A similar result can be shown for virtual storage functions.

THEOREM 7. *If ϕ is a virtual storage function for a cyclodissipative system, then*

$$\phi_a^*(x) \leq \phi(x) \leq \phi_r(x)$$

for all $x \in X$.

PROOF. The upper bound follows from exactly the same argument as in the previous theorem. For the lower bound, consider the collection of trajectories going from $x(t_0) = x_0$ to $x(t_1) = x_1 = 0$. We have

$$\phi(x_0) + E(x_0, u, y, t_0, t_1) \geq \phi(x_1) = 0$$

so that

$$\phi(x_0) \geq -\inf E(x_0, u, y, t_0, t_1)$$

where the infimum is taken over those u and t_1 meeting the boundary condition $x(t_1) = 0$. The value of the infimum is, of course, ϕ_a^*. □

6. Positive storage functions

A storage function is nonnegative for all values of its argument, but sometimes we would like a slightly stronger property. When is it possible to guarantee that $\phi(x) > 0$ for all $x \neq 0$? The answer turns out to depend on observability.

For simplicity, we shall only cover the case of continuous-time systems with an energy input function of the form

$$E(x_0, u, y, t_0, t_1) = \int_{t_0}^{t_1} w(u(t), y(t))\, dt$$

but the argument for discrete-time systems is almost identical.

It does turn out to be necessary to put a restriction on the class of supply rates that are considered.

ASSUMPTION 1. *The supply rate $w(u, y)$ is such that, for any $y \neq 0$, there exists a choice of $u = k(y)$, with $k(0) = 0$, such that $w(k(y), y) < 0$.*

This assumption will turn out to be important in Chapters 8 and 9, as a definition of the class of "interesting" dissipative systems. Note well that it is an assumption about $w(u, y)$, viewed as a function of two independent variables. This is not the same as $w(u, Gu)$, where y is constrained to be the output of the system under consideration.

Now, consider the trajectory of the system with initial state $x(0) = x_0$ and input $u(t) = k(y(t))$. We have

$$\phi(x_0) + \int_0^{t_1} w(k(y(t)), y(t))\, dt \geq \phi(x(t_1)) \geq 0$$

and therefore

$$\phi(x_0) \geq -\int_0^{t_1} w(k(y(t)), y(t))\, dt$$

By Assumption 1, the right side of this inequality is positive except when $y(t) = 0$ for all $t > 0$ (and therefore $u(t) = 0$ for all $t > 0$). If the system has an observable state space, this can happen only if $x_0 = 0$. We deduce, then, that observability implies that $\phi(x) > 0$ for all $x \neq 0$.

7. Lossless systems

Since cyclodissipativeness and dissipativeness are defined in terms of inequalities, the following definition should come as no surprise.

Definition 11. *A system is cyclolossless if it has a virtual storage function satisfying*

$$\phi(x(t_0)) + E(x(t_0), u, y, t_0, t_1) = \phi(x(t_1))$$

for all $t_1 \geq t_0$, all $x(t_0)$, and all u. It is lossless if in addition $\phi(x) \geq 0$ for all x.

Although the definition requires only that one of the virtual storage functions have this property, the following theorem shows that we need not be concerned about what happens with the others.

Theorem 8. *A (cyclo)lossless system has a unique (virtual) storage function.*

Proof. In the case where $x(t_0) = 0$ we have

$$E(x(t_0), u, y, t_0, t_1) = \phi(x(t_1))$$

and therefore

$$\inf_{u \in U_e, t_0 \leq t_1} E(x(t_0), u, y, t_0, t_1) = \phi(x(t_1))$$

That is, $\phi(x(t_1)) = \phi_r(x(t_1))$. The same argument with different boundary conditions shows that $\phi(x(t_1)) = \phi_a^*(x(t_1))$. This shows that a cyclolossless system can have only one virtual storage function. The result also applies to a lossless system, since we know that the set of all storage functions of a dissipative system is a subset of the set of all its virtual storage functions. □

8. More general energy functions

For all of this chapter, we have used the notation $E(x_0, u, y, t_0, t_1)$ to indicate an input energy function that is either an integral (for continuous-time systems) or a sum (for discrete-time systems). It is interesting to note that none of the proofs require that E be in the form of an integral or sum, or indeed any other sort of inner product. The only property that we have needed is the time separation property

$$E(x_0, u, y, t_0, t_2) = E(x_0, u, y, t_0, t_1) + E(x(t_1), u, y, t_1, t_2)$$

It follows, then, that the results of this chapter — although not, unfortunately, the results of most of the following chapters — remain valid for an extremely general definition of what we mean by an input energy function. That more general approach was used in the Willems approach [**Wil72**] to dissipative systems. It should be noted, though, that that approach took the existence of a storage function as the definition of dissipativeness. We have instead preferred to start with an input-output definition, and from that derive the storage functions.

Attractive though that more general approach might be, the focus in the remainder of this book will be on an input energy term that depends quadratically on u and y, because the quadratic dependence will lead to more explicit results. In particular, we will eventually want to work out ways to calculate the storage functions. The quadratic form will allow us to derive equations that the storage functions must satisfy. A more general form for the input energy would lead to serious computational difficulties.

CHAPTER 4

Passive Systems

1. Overview

It is possible that the development of the last couple of chapters has been a little too abstract for some readers. In this chapter, therefore, we take a step backwards and look at some rather more familiar details. There are no new results in this chapter, unless one counts the definitions of strong passivity. Our goal is simply to review some well-known results, in preparation for showing that dissipativeness is, very largely, just an extension of the passivity concept.

It should be clear, from what has been presented so far, that passivity is the same as $(0, I, 0)$ dissipativeness, or $(0, \frac{1}{2}I, 0)$ dissipativeness if one wants to quibble. By extension, we can give the name cyclopassivity to a system that is $(0, I, 0)$ cyclodissipative. A corresponding extension for the property of $(0, I, 0)$ ultimate virtual dissipativeness would lead to a rather clumsy name, but there does not seem to be a need for a new name for that property.

The results in this chapter will be presented without proof. That is because they are special cases of results to be presented in later chapters. The point of this chapter is not to present new theory, but to point out some well-known results with the promise that we will see extensions of those results in the subsequent development.

Most of the well-known results are, as it happens, for linear systems. One of the underlying themes of this book is that linear systems results can, with very little effort, be extended to large classes of nonlinear systems.

2. Passivity and stability

We usually define passivity of a system G via the inequality

$$\langle u, Gu \rangle_T \geq 0 \tag{1}$$

for all $u \in U_e$ and all $T > 0$. For linear systems, it is known that all of the poles of a passive system are in the closed left half-plane. We are therefore inclined to believe that passive systems are stable.

This assertion has to be qualified in several ways. First, the definition above refers to an input-output property. Whenever we refer to the relationship between an input-output property and an internal property, we have to remain aware that a system can have internal hidden modes that do not show up in the input-output description. To be able to assert a connection between internal and external properties, we must require that all states be both reachable and observable. This is well known, but it still needs to be mentioned.

Next, we need to remember that a linear passive system can have poles on the imaginary axis. In such a case, the passive system is still stable in the sense of Lyapunov, but not asymptotically stable. Consequently, it is not input-output stable by any of the usual definitions of input-output stability. For nonlinear systems we cannot talk of poles, but the same general conclusions apply.

With this consideration in mind, we have to alter our stability assertion. It is more accurate to say that a *strongly passive* system is stable, both in an input-output sense and (given minimality, so that there are no hidden states) in the state-space sense that all states are asymptotically stable in the sense of Lyapunov. Precisely what we mean by "strongly" will be clarified towards the end of this chapter.

Another well-known result is that a feedback loop formed by two passive systems is stable. Again, that statement is not quite correct. Stability again requires some form of strong passivity, and this will be covered later in this chapter. For linear systems, the "strong" condition is all about avoiding a phase shift of exactly 90 degrees. For nonlinear systems the notion of "phase shift" is not well-defined, but we will still be able to formulate a suitable definition of strong passivity.

3. Internal and external passivity

External passivity of a system $y = Gu$ is defined by equation 1. Internal passivity is defined in terms of an internal storage function. For linear systems, that storage function is a quadratic $x^T Px$, where x is the state. There is a well-known result called the Positive Real Lemma, or Kalman-Yacubovich-Popov (KYP) Lemma, which can be stated as follows. It says that a realisation of a system, assumed to be completely controllable and completely observable, with state equations

$$\begin{aligned} \frac{dx}{dt} &= Fx + Gu \\ y &= H^T x + Ju \end{aligned}$$

is passive, with storage function $x^T Px$, if and only if there exists a solution $P = P^T > 0$ to the inequality

$$\begin{bmatrix} -PF - F^T P & H - PG \\ H^T - G^T P & J + J^T \end{bmatrix} = \begin{bmatrix} L \\ W^T \end{bmatrix} \begin{bmatrix} L^T & W \end{bmatrix} \geq 0 \qquad (2)$$

If $J + J^T$ is nonsingular, then some (but not all) of the solutions to this inequality can be obtained by setting $W = \left(J + J^T\right)^{1/2}$, which leads to the equation

$$PF + F^T P + (H - PG)\left(J + J^T\right)^{-1}(H - PG)^T = 0$$

It is known that this equation has non-unique solutions. (And that it has no solution $P > 0$ if the system is not passive.) This, however, is only part of the story. It is also known that the set of all solutions P forms a convex set, and the equation just stated gives us only the "corner" points of that set. Equation 2 fixes the number of rows of L and the number of columns of W, but it permits solutions with non-square matrices W. Taking that into consideration, we get a whole continuum of solutions.

Although this derivation considers only the case where $J + J^T$ is nonsingular, similar comments apply where this matrix is singular. We still get a continuum of solutions. Even in the case where $J + J^T = 0$, when the equations reduce to

$$\begin{aligned} PF + F^T P &\leq 0 \\ PG &= H \end{aligned}$$

we have the possibility of multiple solutions.

The preceding discussion applies only to linear systems. A similar result can be obtained for nonlinear systems whose state equations are linear in the control [**Moy74**]. Whether this can be extended to an even broader class of systems is as yet unknown.

4. Passivity in the frequency domain

For a linear system with transfer function $G(s)$, it is natural to use an L_2 inner product, so that the condition

$$\langle u, Gu \rangle_T \geq 0$$

takes the form

$$\int_0^T y(t)^T u(t) dt \geq 0$$

It is then tempting to use Parseval's Theorem to translate this into the frequency domain condition

$$U(-j\omega)^T \left(G(j\omega) + G(-j\omega)^T \right) U(j\omega) \geq 0$$

Strictly speaking, this is an abuse of Parseval's Theorem, and one has to be a little cautious about translating time domain inequalities into the "obvious" frequency domain inequalities. In the present case, however, the translation works, and the correct frequency domain condition is

$$G(s) + G(s)^* \geq 0$$

for all s in Re $s \geq 0$. As usual, the superscript star means complex conjugate transpose. This is a condition that must be satisfied over the entire right half of the complex plane, so is not easy to check. Luckily, this can be simplified down to

$$G(j\omega) + G(-j\omega)^T \geq 0$$

for all real ω, together with the side conditions

- there are no poles in Re $s > 0$;
- for poles on Re $s = 0$, we must satisfy a residue condition that is based on looking at what happens to $G(s)$ as s approaches the pole from the right.

To see where these side constraints come from, suppose that $G(s)$ has a simple pole at $s = s_0$. To keep the explanation simple, let us consider only the case of scalar $G(s)$. In the vicinity of the pole we can say, to a good approximation,

$$G(s) \simeq \frac{a}{s - s_0}$$

where a is a constant. Writing $s - s_0 = \rho e^{j\theta}$, we have

$$G(s) + G(s)^* \simeq \frac{a}{\rho} \left(e^{j\theta} + e^{-j\theta} \right) = \frac{2a}{\rho} \cos\theta$$

This means that, regardless of the sign of a, the result takes both positive and negative values as θ varies. If the pole is on the imaginary axis, the condition $a > 0$ ensures that the negative values occur only in the left half plane. If the pole is in the open right half plane, we have a violation of the frequency domain inequality.

This analysis was for a simple pole. For multiple poles there is an even greater sign fluctuation as θ varies, so multiple poles are acceptable only in the open left half plane.

This shows that those side conditions are necessary. To show that they are also sufficient to ensure that $G(s) + G(s)^* \geq 0$ in the right half plane takes a little more work, but we can omit the proof here because it turns out to be a special case of the results in Chapter 8.

The fact that arguments based on a naive application of Parseval's Theorem are not necessarily valid cannot be overemphasised. For passive systems, this argument works. For more general dissipative systems, there are traps for the unwary.

For nonlinear systems, there is of course no way to express the passivity condition in the frequency domain. Here we are confined to time-domain definitions.

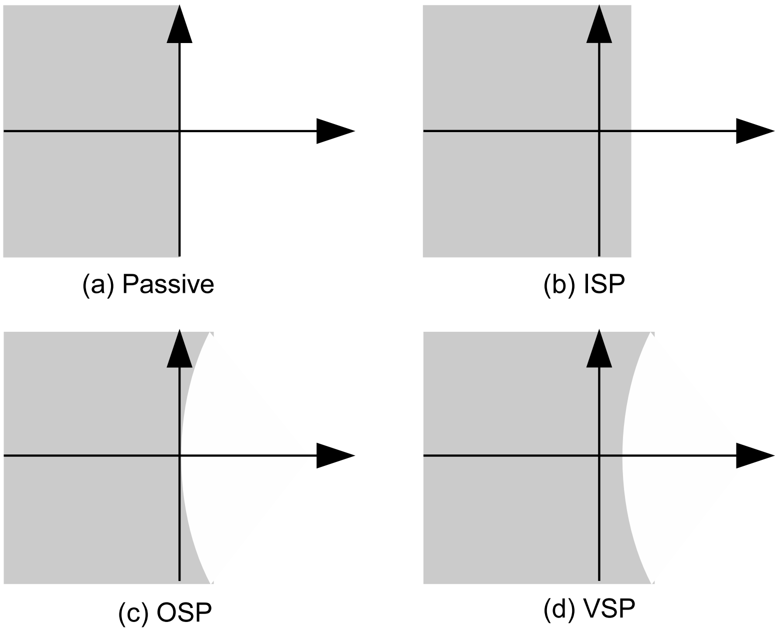

FIGURE 1. Constraints on $\mathbf{G}(j\omega)$ for different kinds of strong passivity

5. Strong passivity

Quite a few system theoretic results that use passivity ideas need a property that is slightly stronger than passivity, and this has led researchers to coin the term "strict passivity". Unfortunately, the variety of results in this area has given rise to a number of non-equivalent definitions of strict passivity. In this section, an attempt will be made to reduce the confusion.

Starting from the passivity definition

$$\langle u, y \rangle_T \geq 0$$

an obvious strengthening of the condition is to require

$$\langle u, y \rangle_T > 0$$

Regrettably, this does not work; the left side of this inequality is zero whenever u or y is zero. The only way to retrieve the situation is to qualify the inequality with a condition like "whenever $u \neq 0$ or $y \neq 0$". Almost equivalently, we must replace the zero bound with a bound that depends on u and/or y.

For linear systems, if we impose a condition of the form

$$\langle u, y \rangle_T \geq \alpha(u, y)$$

then there is absolutely no loss of generality in making the function α quadratic in its arguments. This is not quite true for arbitrary nonlinear systems; but, since we know that arguments about boundedness and related matters often reduce to finding a bound on a norm, it makes sense to adopt a quadratic bound in every case. The remaining question is whether the bound should depend on u, or on y, or both. This leads to the following definitions.

DEFINITION 12. *System G is output strongly passive (OSP) if it is $(-\varepsilon_1 I, I, 0)$ dissipative for some $\varepsilon_1 > 0$. That is, if $2\langle u, y \rangle_T \geq \varepsilon_1 \|y\|_T^2$ for all u.*

DEFINITION 13. *System G is input strongly passive (ISP) if it is $(0, I, -\varepsilon_2 I)$ dissipative for some $\varepsilon_2 > 0$. That is, if $2\langle u, y\rangle_T \geq \varepsilon_2 \|u\|_T^2$ for all u.*

DEFINITION 14. *System G is very strongly passive (VSP) if for some $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ it is $(-\varepsilon_1 I, I, -\varepsilon_2 I)$ dissipative. That is, if $2\langle u, y\rangle_T \geq \varepsilon_1 \|y\|_T^2 + \varepsilon_2 \|u\|_T^2$ for all u.*

For *linear* systems, we can express these constraints in terms of bounds on the transfer function $G(j\omega)$. Figure 1 shows the regions on a Nyquist plot that the transfer function must not enter, for the four different kinds of passivity. The axes are the real and imaginary parts, respectively, of the transfer function, and the shaded regions are the “forbidden” regions where the graph of the transfer function is not allowed to go.

A system which is ISP and also has finite gain is VSP. To see this, notice that the inequalities

$$2\langle u, y\rangle_T \geq \varepsilon_2 \|u\|_T^2$$

and

$$\|y\|_T^2 \leq k^2 \|u\|_T^2$$

can be combined to give

$$2\langle u, y\rangle_T + \alpha k^2 \|u\|_T^2 \geq \varepsilon_2 \|u\|_T^2 + \alpha \|y\|_T^2$$

for any constant $\alpha \geq 0$. After rearranging, this gives

$$2\langle u, y\rangle_T \geq \alpha \|y\|_T^2 + \left(\varepsilon_2 - \alpha k^2\right) \|u\|_T^2$$

which implies the VSP condition provided that we choose α to be in the range $0 < \alpha < \varepsilon_2/k^2$.

On occasion, it is convenient to postulate a property that is slightly weaker than passivity. We can call a system *pseudo strongly passive* if it is $(\varepsilon_1 I, I, \varepsilon_2 I)$ dissipative, with no constraints on the signs of ε_1 and ε_2. (Of course, the values of ε_1 and ε_2 now need to be specified rather than left arbitrary.) Most of the time, however, it is good enough simply to give the dissipativeness parameters rather than to give a special name to this case.

Another definition of strict passivity that is sometimes encountered is the condition: a system with transfer function $G(s)$ is called strictly passive if $G(s-\varepsilon)$, for some $\varepsilon > 0$, is the transfer function of a passive system. This does not fall into any of the classes already discussed. This property can be used in proving exponential stability. For a linear system with input u, output y, and state $x(t)$, consider the change of variables $x_1(t) = e^{\varepsilon t}x(t)$, $u_1(t) = e^{\varepsilon t}u(t)$, and $y_1(t) = e^{\varepsilon t}y(t)$. If the transfer function from u to y is $G(s)$, then it is easy to show that the transfer function from u_1 to y_1 is $G(s-\varepsilon)$. Now, if it is possible to show that $x_1(t)$ is bounded, it obviously follows that $x(t)$ is exponentially bounded.

Attractive as this approach might seem, its usefulness is confined mainly to linear systems. If we try to apply a similar transformation to a time-invariant nonlinear system, the transformed system is time-varying. This immediately increases the difficulty of proving anything interesting about the system. The technique does work sometimes, but perhaps not often enough to be worth pursuing. In any case, we shall hitherto ignore this form of strict passivity, if only because it leads to a complicated definition if we are not able to use transfer functions.

6. The single-loop stability result

One of the best-known results for the stability of a single-loop feedback system is the one that states that if both subsystems are passive, the overall system is stable. That, however, is an oversimplification. In fact we need some sort of strong passivity for the subsystems.

Let us therefore suppose that subsystem i is $(-\varepsilon_{1i}I, I, -\varepsilon_{2i}I)$ dissipative, for $i = 1, 2$. That is, we have

$$\begin{aligned} -\varepsilon_{11}\langle y_1, y_1\rangle_T + 2\langle y_1, u_1\rangle_T - \varepsilon_{21}\langle u_1, u_1\rangle_T &\geq 0 \\ -\varepsilon_{12}\langle y_2, y_2\rangle_T + 2\langle y_2, u_2\rangle_T - \varepsilon_{22}\langle u_2, u_2\rangle_T &\geq 0 \end{aligned}$$

The interconnection equations are

$$\begin{aligned} u_1 &= u_{e1} - y_2 \\ u_2 &= u_{e2} + y_1 \end{aligned}$$

where u_{e1} and u_{e2} are the external inputs. Adding the two inequalities, and substituting the expressions for u_1 and u_2, we get

$$\begin{aligned} &\langle y_1, u_{e1} + \varepsilon_{22}u_{e2}\rangle_T + 2\langle y_2, u_{e2} + \varepsilon_{21}u\rangle_T \\ \geq\ & (\varepsilon_{11} + \varepsilon_{22})\|y_1\|_T^2 + (\varepsilon_{12} + \varepsilon_{21})\|y_2\|_T^2 + \varepsilon_{21}\|u_{e1}\|_T^2 + \varepsilon_{22}\|u_{e2}\|_T^2 \end{aligned}$$

Now define

$$\begin{aligned} k_1 &= \frac{1}{\varepsilon_{11} + \varepsilon_{22}} \\ k_2 &= \frac{1}{\varepsilon_{12} + \varepsilon_{21}} \\ v_1 &= u_{e1} + \varepsilon_{22}u_{e2} \\ v_2 &= u_{e2} + \varepsilon_{21}u_{e1} \end{aligned}$$

After a short calculation, we obtain

$$\frac{1}{k_1}\|y_1 - k_1v_1\|_T^2 + \frac{1}{k_2}\|y_2 - k_2v_2\|_T^2 \leq k_1\|v_1\|_T^2 + k_2\|v_2\|_T^2$$

From here it is a short step to proving finite-gain bounds on the output, provided that k_1 and k_2 are both positive. The conclusion is that the system is stable under any of the following conditions.

- One subsystem is at least passive, and the other is very strongly passive; or
- both subsystems are input strongly passive; or
- both subsystems are output strongly passive.

The above calculations were for input-output stability. For Lyapunov stability, we have the storage function inequalities

$$\begin{aligned} \phi_1(x_1(0)) + 2\langle y_1, u_1\rangle_T - \varepsilon_{11}\|y_1\|_T^2 - \varepsilon_{12}\|u_1\|_T^2 &\geq \phi_1(x_1(T)) \\ \phi_2(x_2(0)) + 2\langle y_2, u_2\rangle_T - \varepsilon_{21}\|y_2\|_T^2 - \varepsilon_{22}\|u_2\|_T^2 &\geq \phi_2(x_2(T)) \end{aligned}$$

where $x_1(t)$ and $x_2(t)$ are the states of the two subsystems. With zero external input, we have $u_1 = -y_2$, $u_2 = y_1$, and therefore

$$\begin{aligned} \phi_1(x_1(0)) - 2\langle y_1, y_2\rangle_T - \varepsilon_{11}\|y_1\|_T^2 - \varepsilon_{12}\|y_2\|_T^2 &\geq \phi_1(x_1(T)) \\ \phi_2(x_2(0)) + 2\langle y_2, y_1\rangle_T - \varepsilon_{21}\|y_2\|_T^2 - \varepsilon_{22}\|y_1\|_T^2 &\geq \phi_2(x_2(T)) \end{aligned}$$

Now let $\phi(x) = \phi(x_1, x_2) = \phi_1(x_1) + \phi_2(x_2)$. By adding the two inequalities, we get

$$\phi(x(0)) - (\varepsilon_{11} + \varepsilon_{22})\|y_1\|_T^2 - (\varepsilon_{12} + \varepsilon_{21})\|y_2\|_T^2 \geq \phi(x(T))$$

From this is should be clear that $\phi(x(t))$ is monotonically non-increasing, provided that $(\varepsilon_{11} + \varepsilon_{22})$ and $(\varepsilon_{12} + \varepsilon_{21})$ are both positive. This means that $\phi(x)$ will work as a Lyapunov function if the overall system has the property that $y_1 \to 0$ and $y_2 \to 0$ together imply that $x \to 0$. This property can be guaranteed with an observability assumption on each subsystem separately.

CHAPTER 5

Algebraic conditions for dissipativeness

1. Overview

Until now we have not looked at the question of how we can decide whether a given system is (Q, S, R) dissipative. In the present chapter we are going to focus on a special class of nonlinear systems: those whose state equations are linear in the input. It will be shown that, for these systems, we can write down equations for the (virtual) storage function of a cyclodissipative or dissipative system. That means that the question of cyclodissipativeness and dissipativeness reduces down to whether solutions exist to these equations.

The systems in question include linear systems as a special case. In that case the question becomes one of being able to solve a matrix Riccati equation.

We shall also look briefly at discrete-time systems. For linear systems we get a similar set of matrix equations. For nonlinear discrete-time systems, unfortunately, it does not appear possible to obtain simple criteria.

2. A class of nonlinear continuous-time systems

For most of this chapter, we will be concerned with systems with state equations

$$\begin{aligned} \frac{dx}{dt} &= f(x(t)) + G(x(t))u(t) \\ y(t) &= h(x(t)) + J(x(t))u(t) \end{aligned} \tag{3}$$

Note that the state equations are linear in the input $u(t)$. The reason for concentrating on this class of systems is that it will turn out to be possible to get explicit conditions for dissipativeness and cyclodissipativeness, in terms of a set of equations that the storage function or virtual storage function must satisfy.

Our starting point is the inequality

$$\phi(x(t_0)) + \int_{t_0}^{t_1} \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t)\right) dt \geq \phi\left(x\left(t_1\right)\right)$$

which can be put into differential form

$$y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t) \geq \lim_{\delta t \to 0} \frac{\phi\left(x\left(t+\delta t\right)\right) - \phi(x(t))}{\delta t}$$

At this point, strictly speaking, we should be looking for conditions under which $\phi(x(t))$ is differentiable, but that would take us into technicalities that would obscure the main line of the argument. Let us therefore assume, for now, that the differentiability question has been settled. Then we get

$$\frac{d}{dt}\phi(x(t)) = \frac{d\phi}{dx}^T \frac{dx}{dt} = \nabla\phi^T \left(f(x(t)) + G(x(t))u(t)\right)$$

where we are using the notation $\nabla\phi$ to denote the derivative of $\phi(x)$ with respect to x. (As distinct from the derivative of $\phi(x(t))$ with respect to t.) Expanding out the inequality, we get

$$(h + Ju)^T Q(h + Ju) + 2(h + Ju)^T S u + u^T R u \geq \nabla\phi^T f + \nabla\phi^T G u$$

where, for the sake of readability, we have dropped the independent variables from the notation. Expanding this one further step, we have

$$\begin{aligned} &h^T Q h + 2h^T Q J u + u^T J^T Q J u \\ &\qquad +2h^T S u + 2u^T J^T S u + u^T R u - \nabla\phi^T f - \nabla\phi^T G u \geq 0 \end{aligned}$$

This is an inequality that must hold pointwise for all x, and all u. Note that we have reached the point in this calculation where the quantity on the left may be considered as a function of x and u; we have, in effect, frozen the calculations at one instant of time. We no longer have to consider the time evolution of these variables. The variable x may be considered as being the initial state for this point in time, and because of this we can treat x and u as independent variables.

Now, although the quantity on the left of the inequality might be a complicated nonlinear function of x, it contains only linear and quadratic functions of u. (This is because we have restricted our state equations to be linear in u.) A quadratic function that is always nonnegative can always be put into the form of a perfect square plus a constant, although in this case our "constant" will depend on the other variable x. That is, we have a function that can be expressed as

$$(u + a(x))^T B(x) (u + a(x)) + c(x)$$

for some a, B, and c whose precise form has not yet been determined, except that we know that B must be nonnegative definite and the scalar c must be nonnegative. It will actually be more convenient for us to write this quadratic form as

$$(\ell(x) + W(x)u)^T (\ell(x) + W(x)u)$$

which involves no loss of generality, since the two forms can be shown to be equivalent. One way to make the expressions match up is to factor B as $W_1^T W_1$, define $\ell_1 = W_1 a$, factor c as $\ell_2^T \ell_2$, and set

$$\ell = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix} \text{ and } W = \begin{bmatrix} W_1 \\ 0 \end{bmatrix}$$

Note, however, that this is just one of many possibilities. The choice of ℓ and W is non-unique.

Putting all of this together, we get

$$\begin{aligned} &h^T Q h + 2h^T Q J u + u^T J^T Q J u + 2h^T S u \\ &\qquad +2u^T J^T S u + u^T R u - \nabla\phi^T f - \nabla\phi^T G u = (\ell + W u)^T (\ell + W u) \end{aligned}$$

Equating coefficients of u, this reduces down to the three equations

$$\begin{aligned} h^T Q h - \nabla\phi^T f &= \ell^T \ell \\ h^T Q J + h^T S - \frac{1}{2}\nabla\phi^T G &= \ell^T W \\ J^T Q J + J^T S + S^T J + R &= W^T W \end{aligned}$$

3. The main result for continuous-time systems

We are now in a position to state the main result of this chapter. It is actually two results, depending on whether we are talking about dissipativeness or cyclodissipativeness.

THEOREM 9. *The necessary and sufficient condition for a differentiable function $\phi(x)$ to be a virtual storage function for (Q, S, R) cyclodissipativeness for the* *system with state equations 3 is that there exist functions $\ell(x)$ and $W(x)$ satisfying the equations*

$$\begin{aligned}\nabla\phi^T f &= h^T Q h - \ell^T \ell \\ \frac{1}{2}\nabla\phi^T G &= h^T(QJ+S) - \ell^T W \\ R + J^T S + S^T J + J^T Q J &= W^T W\end{aligned}$$

PROOF. Necessity was established in the last section, by showing that the virtual storage inequality can be reduced down to these equations. For sufficiency, we simply reverse the argument. Observe that

$$\begin{aligned}\phi(x(t_1)) &= \phi(x(t_0)) + \int_{t_0}^{t_1} \frac{d\phi(x(t))}{dt} dt \\ &= \phi(x(t_0)) + \int_{t_0}^{t_1} \nabla\phi^T (f + Gu)\, dt\end{aligned}$$

and the expansion from this point onwards is obvious. □

The corresponding result for dissipativeness is, of course, only a minor modification of the above.

THEOREM 10. *The necessary and sufficient conditions for a differentiable function $\phi(x)$ to be a storage function for (Q, S, R) dissipativeness for the system with state equations 3 are that $\phi(0) = 0$, $\phi(x) \geq 0$ for all x, and there exist functions $\ell(x)$ and $W(x)$ satisfying the equations*

$$\begin{aligned}\nabla\phi^T f &= h^T Q h - \ell^T \ell \\ \frac{1}{2}\nabla\phi^T G &= h^T(QJ+S) - \ell^T W \\ R + J^T S + S^T J + J^T Q J &= W^T W\end{aligned}$$

PROOF. Identical to the proof of the previous theorem, except for the extra condition $\phi(x) \geq 0$. □

Observe that the above was derived by turning an inequality into an equality, by introducing the new quantities $\ell(x)$ and $W(x)$. Taking the calculation in the reverse direction, we get

$$\begin{aligned}\frac{d\phi(x(t))}{dt} &= \nabla\phi^T f + \nabla\phi^T G u \\ &= h^T Q h - \ell^T \ell + 2h^T(QJ+S)u - 2\ell^T W u\end{aligned}$$

which finally reduces down to

$$\frac{d\phi(x(t))}{dt} = y^T Q y + 2y^T S u + u^T R u - (\ell + Wu)^T(\ell + Wu)$$

Integrating this, we get

$$\begin{aligned}\phi(x(t_1)) = \phi(x(t_0)) + \int_{t_0}^{t_1} &\left(y^T Q y + 2y^T S u + u^T R u\right) dt \\ &- \int_{t_0}^{t_1} (\ell + Wu)^T(\ell + Wu) dt\end{aligned}$$

This is, of course, the inequality defining dissipativeness, except that we have now explicity identified the dissipation terms.

4. The question of differentiability

Our results so far leave open the possibility of non-differentiable (virtual) storage functions. In general, it is possible that a dissipative system has some differentiable storage functions and some non-differentiable ones. For our purposes, though, it will be good enough if we can find just one differentiable storage function.

One way to tackle this problem is to impose a condition of local controllability, as defined in [**HM80c**].

Definition 15. *A dynamical system is said to be locally controllable at x_0, with respect to supply rate $w(x,u)$, if for any x_1 in a suitable small open neighbourhood Ω of x_0 there exist choices of $u \in U_e$ and t_1 such that the state can be driven from $x(t_0) = x_0$ to $x(t_1) = x_1$ with the constraint*

$$\left| \int_{t_0}^{t_1} w(x(t), u(t) dt \right| \leq \rho\left(\|x_1 - x_0\|\right)$$

for some continuous function $\rho : R_+ \to R_+$ such that $\rho(0) = 0$. The dynamical system is said to be locally controllable if it is locally controllable at every state $x_0 \in X$.

With this definition we have the following result.

Theorem 11. *Let a dynamical system be locally controllable. Then any virtual storage function that exists for all $x \in X$ is also continuous.*

Proof. Consider some arbitrary state x_0 in X and let the virtual storage function be $\phi(\cdot)$. Then for any x_1 in the neighbourhood Ω of x_0, we have

$$\phi(x_0) + \int_{t_0}^{t_1} w(x(t), u(t) dt \geq \phi(x_1)$$

for the u and the t_1 specified in the definition of local controllability. Considering transitions in each direction between x_0 and x_1, it is easy to see that

$$|\phi(x_1) - \phi(x_0)| \leq \rho\left(\|x_1 - x_0\|\right)$$

which is sufficient to prove continuity of $\phi(\cdot)$. □

Note that this proves only continuity, not differentiability. That, however, is sufficient to allow the earlier results to go through, provided that we define the time derivative of $\phi(x(t))$ along state trajectories as

$$\frac{d\phi(x(t))}{dt} = \limsup_{h \to 0+} \frac{1}{h}\left(\phi(x(t+h)) - \phi(x(t))\right)$$

5. More general nonlinear systems

For completeness, let us note that if the state equations are of the form

$$\begin{aligned} \frac{dx}{dt} &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \end{aligned}$$

then the differential form of the dissipation inequality becomes

$$h^T Q h + 2h^T S u + u^T R u - \nabla\phi^T f \geq 0$$

The difficulty in that case is, because the functions f and h depend on both x and u, we cannot separate out the terms that are quadratic in u and those that are linear in u. In other words, the above inequality does not simplify any further.

6. Linear continuous-time systems

In this section we consider linear time-invariant state equations of the form

$$\begin{aligned} \frac{dx}{dt} &= Fx + Gu \\ y(t) &= H^T x + Ju \end{aligned} \tag{4}$$

For linear systems, the determination of the available storage and required supply are linear-quadratic optimisation problems, and it is known that the optimal value of the performance index for such problems is quadratic in the state. That means that we can put $\phi(x) = x^T P x$ into the equations that the (virtual) storage functions must satisfy, to get

$$\begin{aligned} 2x^T PFx &= x^T HQH^T x - \ell(x)^T \ell(x) \\ x^T PG &= x^T H(QJ + S) - \ell(x)^T W(x) \\ R + J^T S + S^T J + J^T QJ &= W(x)^T W(x) \end{aligned}$$

Obviously this requires that $W(x)$ be a constant matrix, and that $\ell(x)$ be a linear function of x. That means that the equations further simplify down to the equations

$$\begin{aligned} PF + F^T P &= HQH^T - LL^T \\ PG &= H(QJ + S) - LW \\ R + J^T S + S^T J + J^T QJ &= W^T W \end{aligned}$$

This can also, if desired, be put in the form of an inequality by eliminating L and W. The result in that case is

$$\begin{bmatrix} HQH^T - PF - F^T P & H(QJ + S) - PG \\ (QJ + S)^T H^T - G^T P & R + J^T S + S^T J + J^T QJ \end{bmatrix} = \begin{bmatrix} L \\ W^T \end{bmatrix} \begin{bmatrix} L^T & W \end{bmatrix} \geq 0$$

Note that the number of rows of L is equal to the number of states, and the number of columns of W is equal to the number of inputs. On the other hand, the number of rows of W — and therefore the number of columns of L — remains unspecified, and this is one reason why the solution for P is non-unique. Even if we choose a W with the minimum number of rows, the solution is still not unique. To see this, consider the simplest case, where $\left(R + J^T S + S^T J + J^T QJ\right)$ is nonsingular. In that case an obvious choice of W is

$$W = \left(R + J^T S + S^T J + J^T QJ\right)^{1/2}$$

after which it is clear how to eliminate L from the equations. The result is then a matrix Riccati equation, and it is well-known that such equations have multiple solutions in general. The algebra becomes a little trickier when W cannot be written as the square root of a nonsingular matrix, but again the result is a Riccati equation.

7. Linear discrete-time systems

The emphasis so far has been on continuous-time systems, but the theory is also applicable to discrete-time systems. Let us therefore consider systems of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{5}$$

where we have chosen (A, B, C, D) as the notation for the coefficient matrices, rather than (F, G, H, J), simply because this tends to be the traditional choice in the literature. A suitable truncated inner product in this case is

$$\langle f, g\rangle_T = \sum_{t=t_0}^{T-1} f(t)^T g(t)$$

That means that the dissipativeness inequality becomes

$$\phi(x(t_0)) + \sum_{t=t_0}^{t_1-1} \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t)\right) \geq \phi\left(x\left(t_1\right)\right)$$

For a discrete-time system, of course, the question of a time derivative does not arise. Instead we want to look at a single time step, giving the result

$$\phi(x(t)) + y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t) \geq \phi\left(x\left(t+1\right)\right)$$

To short-cut a tedious calculation, let us assume from the outset that the storage function is a quadratic, $x^T P x$. Expanding out the terms, we come eventually to

$$\begin{array}{c} x^T(P + C^TQC - A^TPA)x + 2x^T(C^TQD + C^TS - A^TPB)u \\ +u^T(R + D^TS + S^TD + D^TQD - B^TPB)u \geq 0 \end{array}$$

This can be written as

$$\left[\begin{array}{cc} x^T & u^T \end{array}\right] M \left[\begin{array}{c} x \\ u \end{array}\right] \geq 0$$

where the coefficient matrix is

$$M = \left[\begin{array}{cc} P + C^TQC - A^TPA & C^TQD + C^TS - A^TPB \\ D^TQC + S^TC - B^TPA & R + D^TS + S^TD + D^TQD - B^TPB \end{array}\right]$$

Since this must be true for all x and all u, the coefficient matrix M is nonnegative definite, and may be factored as

$$M = \left[\begin{array}{c} L \\ W^T \end{array}\right] \left[\begin{array}{cc} L^T & W \end{array}\right]$$

The dissipativeness or cyclodissipativeness condition therefore reduces to requiring a solution of the equations

$$\begin{array}{rcl} P + C^TQC - A^TPA & = & LL^T \\ C^TQD + C^TS - A^TPB & = & LW \\ R + D^TS + S^TD + D^TQD - B^TPB & = & W^TW \end{array}$$

If these equations have a solution P, then the system is (Q, S, R) dissipative or cyclodissipative. For dissipativeness, P has to be nonnegative definite. For cyclodissipativeness, we do not care about the sign of P.

A similar approach does not, unfortunately, work for the obvious class of nonlinear systems. This is because the expansion of $\phi(Ax + Bu)$ in the calculations relies on the fact that $\phi(\cdot)$ is quadratic. For nonlinear systems with non-quadratic storage functions, no comparable results are known.

CHAPTER 6

Stability

1. Overview

Having set the scene, we now wish to present what are arguably the most important results of this book. Our basic result is that a (Q, S, R) dissipative system is stable if $Q < 0$. That does not mean that the case $Q \geq 0$ is not interesting. We are not able to deduce stability in such cases, but it will still often turn out to be possible to prove stability of a larger system of which the $Q \geq 0$ system is a subsystem.

The basic stability result is, in practice, not as interesting as it might seem. In later chapters, we will be looking at how to find the (Q, S, R) triples for which a given system is (Q, S, R) dissipative. It will turn out that it is easiest to do this in two steps. First, find a (Q, S, R) triple for which the system is (Q, S, R) cyclo-dissipative. Next, find side conditions that guarantee that the cyclo-dissipative system is, in fact, dissipative. In the case $Q < 0$, the side condition turns out to be that the system is stable. It appears, then, that we need to know that the system is stable in order to prove that it is stable.

Luckily, all is not lost. We probably do not need stability criteria for simple systems anyway. If a system is simple enough, we can use *ad hoc* methods to check its stability. Where we do need powerful tools is in the case of complex systems.

That is what most of this chapter is about. We consider a complex system to be an interconnection of simpler systems. Sometimes that decomposition is an obvious consequence of seeing that the system is built from subsystems. Sometimes it is because we put an arbitrary subdivision between the equations that describe the system. It does not matter. All that matters is that we have an "interconnection" model of our complex system.

Next, we try to find dissipativeness parameters for each subsystem. We do not show how to do that in this chapter, but it will be covered in later chapters.

Finally, we combine those dissipativeness parameters, in such a way that we end up with a sufficient condition for stability. Not a necessary and sufficient condition, it is true, but we cannot have everything. In the next chapter we will look at some sufficient conditions for instability, and that will go at least part of the way towards closing the gap.

2. The basic stability results

We actually need two basic stability results. One for input-output stability, and one for state-space stability. Let us begin with the input-output result.

THEOREM 12. *If system G is (Q, S, R) dissipative for some $Q < 0$, then G is finite-gain stable.*

PROOF. We have

$$\langle y, Qy \rangle_T + 2 \langle y, Su \rangle_T + \langle u, Ru \rangle_T \geq 0$$

for all $u \in U_e$ and all T, where $y = Gu$. Since Q is self-adjoint and negative definite, we can factor it as $Q = -M^* M$ where M is invertible. Let $K = (M^*)^{-1} S$. Then we have

$$
\begin{aligned}
&\|My - Ku\|_T^2 \\
&\quad = \langle My - Ku, My - Ku\rangle_T \\
&\quad = \langle y, M^*My\rangle_T - \langle y, M^*Ku\rangle_T - \langle M^*Ku, y\rangle_T + \langle u, K^*Ku\rangle_T \\
&\quad = -\langle y, Qy\rangle_T - 2\langle y, Su\rangle_T - \left\langle u, S^TQ^{-1}Su\right\rangle_T \\
&\quad = -\langle y, Qy\rangle_T - 2\langle y, Su\rangle_T - \langle u, Ru\rangle_T + \left\langle u, \left(R - S^TQ^{-1}S\right)u\right\rangle_T \\
&\quad \leq \left\langle u, \left(R - S^TQ^{-1}S\right)u\right\rangle_T
\end{aligned}
$$

Now, let a scalar $\alpha > 0$ be chosen such that $R - S^TQ^{-1}S \leq \alpha^2 I$. (Obviously, such an α always exists.) This gives

$$\|My - Ku\|_T^2 \leq \alpha^2 \|u\|^2$$

and then

$$\|My\| \leq \|My - Ku\| + \|Ku\| \leq (\alpha + \|K\|)\,\|u\|$$

So finally

$$\|y\| \leq \left\|M^{-1}\right\|(\alpha + \|K\|)\,\|u\|$$

which establishes a gain bound for the system. □

An interesting feature of this proof is the role of the matrix $R - S^TQ^{-1}S$. At various places in this book we have raised the question of the class of "interesting" (Q, S, R) triples. The α of this proof is clearly related to the largest (positive) eigenvalue of $R - S^TQ^{-1}S$. If $R - S^TQ^{-1}S < 0$ then it would appear that the inequality in the proof could never be satisfied. In fact, no system can possibly be (Q, S, R) dissipative if $R - S^TQ^{-1}S < 0$. This simply underlines the fact that not all possible (Q, S, R) triples are meaningful in the definition of dissipativeness. For the meaningless ones the theorem is still true, in a technical sense, but only in the sense that "if pigs could fly, this system would be stable".

For completeness, we should cover the property of weak dissipativeness that was introduced in Chapter 2. Recall that a system described in input-output terms is called weakly (Q, S, R) dissipative if there exists a constant β such that

$$\langle y, Qy\rangle_T + 2\langle y, Su\rangle_T + \langle u, Ru\rangle_T + \beta \geq 0$$

for all $u \in U_e$ and all T. In the special case where $(Q, S, R) = (-k^2 I, 0, I)$ we say that the system is weakly finite-gain stable. For many purposes, weak finite gain stability is a sufficiently good property.

Theorem 13. *If system G is weakly (Q, S, R) dissipative for some $Q < 0$, then G is weakly finite-gain stable.*

Proof. Identical to the proof of Theorem 12. □

Now, let us turn to the question of state-space stability. As you might expect, the corresponding result requires the system to be well-posed in the sense that input-output stability can be related to state-space stability. In fact we require the following property.

Definition 16. *System G is zero-state detectable (ZSD) if there exists some $T > 0$ and a continuous strictly monotonic function $\alpha : R^+ \to R^+$, with $\alpha(0) = 0$ and $\alpha(\sigma) > 0$ for all $\sigma > 0$, such that with $u(t) = 0$ for all $t \geq t_0$, and any $x(t_0) = x_0 \in X$, the output satisfies*

$$\int_{t_0}^{t_0+T} y(t)^T y(t)\,dt \geq \alpha(|x_0|)$$

where $|\cdot|$ is the metric on X.

This is a weak observability condition. For linear systems, it is precisely equivalent to observability.

THEOREM 14. *Suppose that the input-output map G has a ZSD state-space realisation. Then G is asympotically stable if G is (Q,S,R) dissipative for some $Q<0$.*

PROOF. Dissipativeness implies the existence of a storage function ϕ with the property that

$$\phi(x(t_0)) + \int_{t_0}^{t_1} \left(y(t)^TQy(t) + 2y(t)^TSu(t) + u(t)^TRu(t)\right) dt \geq \phi\left(x\left(t_1\right)\right)$$

With zero input, that is $u(t)=0$ for all $t \geq t_0$, this gives

$$\phi(x(t_0)) + \int_{t_0}^{t_1} y(t)^TQy(t)dt \geq \phi\left(x\left(t_1\right)\right)$$

If $Q<0$, this says that $\phi\left(x\left(t\right)\right)$ is monotonically non-increasing with t. Since we have a lower bound $\phi(x) \geq 0$, this says that we must be converging to a final value ϕ_f. For states in this final set, we have

$$\phi_f + \int_{t_0}^{t_1} y(t)^TQy(t)dt \geq \phi_f$$

which can be satisfied only if $y(t)=0$ for all t in the interval. The ZSD condition then implies that $x(t)$ must be converging to 0. □

An alternative proof, which some might prefer, uses the ZSD definition more directly. Note first that

$$y(t)^TQy(t) \leq -\lambda y(t)^Ty(t)$$

where λ is the largest (least negative) eigenvalue of Q. The dissipation inequality and the ZSD definition then imply

$$\lambda\alpha\left(\left|x\left(t\right)\right|\right) \leq \phi\left(x\left(t\right)\right) - \phi\left(x\left(t+T\right)\right)$$

Next, the fact that $\phi\left(x\left(t\right)\right)$ is monotonically approaching a limit from above implies that, given any $\epsilon > 0$, there exists some t such that $\phi\left(x\left(t\right)\right) - \phi\left(x\left(t+T\right)\right) < \epsilon$, independently of T. This tells us that

$$\lambda\alpha\left(\left|x\left(t\right)\right|\right) < \epsilon$$

from which it is clear that $\left(x\left(t\right)\right)$ is converging to zero.

An interesting feature of the above proofs is that we do not initially know the value of the limit ϕ_f. It is only after we have proved that the state tends to zero that we can also conclude that ϕ_f=0.

This completes the proof of our most basic results. Let us now turn to interconnected systems.

3. Interconnected systems

The basic scenario in this section is that we have N subsystems which are linearly interconnected. The "linearly" is not an important restriction. If we have a nonlinear connection, we can always put the nonlinearity into a new subsystem. That is, all the nonlinearity — and, for that matter, all of the dynamic behaviour — is in the subsystems, not in the interconnections.

We suppose that subsystem i is (Q_i, S_i, R_i) dissipative, for $i = 1..N$. We can define an overall input vector u which is a column vector of all the individual subsystem inputs u_i, and similarly for the outputs. Now, let us suppose that all of those subsystems are interconnected via a connection equation

$$u = u_e - Hy \tag{6}$$

where u_e represents the external inputs.

This is a linear interconnection. As explained above, this involves no loss of generality. Any nonlinearities can be put into new subsystems.

An obvious objection at this point is that the dimensionality of u_e is the same as that of u. That is, for every subsystem input, there is a corresponding external input. For typical interconnected systems there are fewer external inputs than internal inputs. That suggests that we should really be working with a model

$$u = Ju_e - Hy$$

where J is a non-square matrix. It turns out, however, that our stability results are independent of J, so there is no point in introducing that generalisation.

To see why, consider a subsystem i, which has an input (because of feedback from other subsystems), but no external input. In real life, there will be noise entering at that input, and the model of that noise will be an external input. If the noise is damped by the other subsystems, so that it has no significant effect on the overall response, then we are justified in ignoring it. If, however, the noise stimulates an unstable mode, then we will have a system that is unstable even if a zero-noise assumption would have ignored that unstable mode. The conclusion, clearly, is that we should include all of those "noise" inputs in a stability analysis; we should not confine our attention to the "deliberate" inputs.

To proceed, then, let us take all of those external inputs into account. For subsystem i we have

$$\langle y_i, Q_i y_i \rangle_T + 2 \langle y_i, S_i u_i \rangle_T + \langle u_i, R_i u_i \rangle_T \geq 0$$

Putting all of those equations together, we get

$$\langle y, Qy \rangle_T + 2 \langle y, Su \rangle_T + \langle u, Ru \rangle_T \geq 0$$

where the three large matrices Q, S, and R are defined as

$$\begin{aligned} Q &= \text{block diag}\{Q_1, ..., Q_N\} \\ S &= \text{block diag}\{S_1, ..., S_N\} \\ R &= \text{block diag}\{R_1, ..., R_N\} \end{aligned}$$

That means that we have

$$\langle y, Qy \rangle_T + 2 \langle y, Su_e - SHy \rangle_T + \langle u_e - Hy, Ru_e - RHy \rangle_T \geq 0$$

or equivalently

$$\left\langle y, \widehat{Q}y \right\rangle_T + 2 \left\langle y, \widehat{S}u_e \right\rangle_T + \left\langle u_e, \widehat{R}u_e \right\rangle_T \geq 0$$

where

$$\begin{aligned} \widehat{Q} &= Q - SH - H^T S^T + H^T RH \\ \widehat{S} &= S - H^T R \\ \widehat{R} &= R \end{aligned}$$

That means that the overall system is $(\widehat{Q}, \widehat{S}, \widehat{R})$ dissipative.

Our overall input-output result is immediate.

THEOREM 15. *The interconnected system defined by equation 6 is finite-gain stable if*

$$\widehat{Q} = Q - SH - H^T S^T + H^T RH < 0$$

If, in addition, each subsystem has a ZSD state-space realisation, then the overall state-space model whose state is the concatenation of the individual state vectors is asymptotically stable.

Proof. The input-output result follows immediately from the above calculations. For the state-space result, we need only check the ZSD condition. For the subsystems, we have

$$\int_{t_0}^{t_0+T} y_i(t)^T y_i(t)dt \geq \alpha_i\left(|x_i\left(t_0\right)|\right)$$

where T is the maximum of the T values in the individual ZSD conditions. Adding these together gives

$$\int_{t_0}^{t_0+T} y(t)^T y(t)dt \geq \sum_i \alpha_i\left(|x_i\left(t_0\right)|\right)$$

Obviously, the function on the right side has the right properties to serve as the α in the ZSD definition. □

A disadvantage of this proof is that it does not show how to construct the overall storage function. Let us therefore take a different approach, working from the storage functions for the subsystems. Our starting point is

$$\phi_i(x_i(t_0)) + \langle y_i, Q_i y_i\rangle_T + 2\langle y_i, S_i u_i\rangle_T + \langle u_i, R_i u_i\rangle_T \geq \phi_i(x_i(t_1))$$

Zero external input means $u = -Hy$, so that $u_i = -\sum H_{ij} y_j$, and we get

$$\begin{aligned}\phi_i(x_i(t_0)) + \langle y_i, Q_i y_i\rangle_T - 2\left\langle y_i, S_i \sum_j H_{ij} y_j\right\rangle_T \\ + \left\langle \sum_j H_{ij} y_j, R_i \sum_j H_{ij} y_j\right\rangle_T \geq \phi_i(x_i(t_1))\end{aligned}$$

We can add these to get

$$\begin{aligned}\sum_i \phi_i(x_i(t_0)) + \sum_i \langle y_i, Q_i y_i\rangle_T - 2\sum_i\left\langle y_i, S_i \sum_j H_{ij} y_j\right\rangle_T \\ + \sum_i\left\langle \sum_j H_{ij} y_j, R_i \sum_j H_{ij} y_j\right\rangle_T \geq \sum_i \phi_i(x_i(t_1))\end{aligned}$$

It is not hard to show that this reduces down to

$$\sum_i \phi_i(x_i(t_0)) + \left\langle y, \widehat{Q} y\right\rangle_T \geq \sum_i \phi_i(x_i(t_1))$$

This suggests that the sum of the subsystem storage functions is a storage function for the overall system, and indeed that turns out to be the case when we repeat the calculation with the external inputs u_e included. (There might, of course, be other storage functions that are not derived that way, but all that matters here is that we have found at least one storage function.) It also shows how to proceed in using that storage function as a Lyapunov-like function for deducing asymptotic stability.

As a matter of clarification, it is worth pointing out that our storage functions are not, strictly speaking, Lyapunov functions, because we have not shown that they are differentiable. The usual route to showing Lyapunov stability is to show that the time derivative of the Lyapunov function along zero-input trajectories is non-positive, and strictly negative for enough of the time. We have chosen not to go down that route, because differentiability or even continuity would require technical assumptions that are not really essential. All that needs to be shown is that the function is monotonically non-increasing, and we can do that without looking at derivatives.

In the rest of this chapter, to avoid tedious repetition, we will use the phrase "the system is stable" to mean both input-output stability and state-space asymptotic stability. In interpreting that phrase, we have to remember that state-space stability requires two extra assumptions, namely (a) that a state-space model exists for each subsystem, and (b) a ZSD assumption for each subsystem. The ZSD assumption basically means that there are no "hidden states" that do not show up in the output.

4. Neutral interconnections

The matrix H is said to describe a *neutral interconnection* if the dissipativeness parameters of the interconnected system are the same as the dissipativeness parameters of the original collection of subsystems. If we recall the equations

$$\begin{aligned}\widehat{Q} &= Q - SH - H^T S^T + H^T RH \\ \widehat{S} &= S - H^T R \\ \widehat{R} &= R\end{aligned}$$

it can be seen that $(\widehat{Q}, \widehat{S}, \widehat{R}) = (Q, S, R)$ if and only if $H^T R = 0$ and $SH + H^T S^T = 0$.

The simplest case of a neutral interconnection is where all the subsystems are connected in parallel. That is, when all subsystems share the same external input, and there is no feedback. In this case $H = 0$, so the conditions are satisfied trivially.

In more typical examples, the interconnection matrix H is likely to have full column rank, so that we can have a neutral interconnection only if $R = 0$.

The most interesting case is where $S = I$ and $R = 0$. In that case, $(\widehat{Q}, \widehat{S}, \widehat{R}) = (Q, I, 0)$ if and only if $H + H^T = 0$. As we shall see in the following section, a feedback connection of two subsystems satisfies this condition. It is well known that a feedback connection of two passive systems is itself passive. It can now be seen that that result extends to the case of $(Q, I, 0)$ dissipative systems, for arbitrary Q.

5. Single-loop feedback systems

To get some feel for what our basic result means, let us look at one of the simplest forms of interconnection: a feedback loop made up of two subsystems.

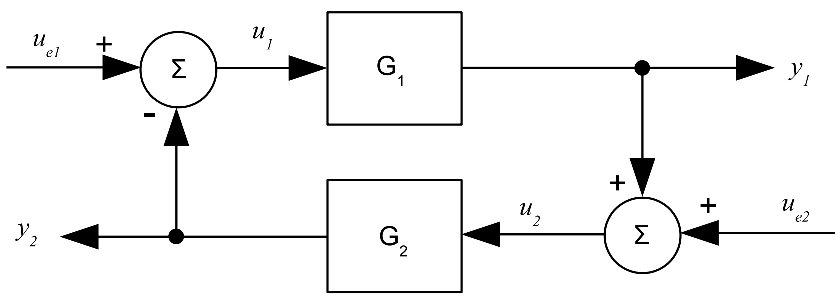

FIGURE 1. A single-loop system

The interconnection equations are

$$\begin{aligned} u_1 &= u_{e1} - y_2 \\ u_2 &= u_{e2} + y_1 \end{aligned}$$

so that the interconnection matrix is

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

This leads to

$$\begin{aligned} \widehat{Q} &= Q - SH - H^T S^T + H^T R H \\ &= \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} - \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} - \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \begin{bmatrix} S_1^T & 0 \\ 0 & S_2^T \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \end{aligned}$$

which finally reduces down to

$$\widehat{Q} = \begin{bmatrix} Q_1 + R_2 & S_2^T - S_1 \\ S_2 - S_1^T & Q_2 + R_1 \end{bmatrix}$$

It can be seen that there is some sort of trade-off between the Q parameter of one subsystem and the R parameter of the other. If, for example, the subsystems are $(-\varepsilon_{11} I, I, -\varepsilon_{21} I)$ dissipative and $(-\varepsilon_{12} I, I, -\varepsilon_{22} I)$ dissipative respectively, then

$$\widehat{Q} = \begin{bmatrix} -(\varepsilon_{11} + \varepsilon_{22}) I & 0 \\ 0 & -(\varepsilon_{12} + \varepsilon_{21}) I \end{bmatrix}$$

From this it is obvious that stability follows if both systems are ISP, or both are OSP, or one is passive and the other is VSP. You can even get stability results when some of the ε_{ij} are negative, provided that the relevant sums are positive.

For another example, suppose that both subsystems have finite gain. This time we have

$$\widehat{Q} = \begin{bmatrix} -1 + k_2^2 & 0 \\ 0 & -1 + k_1^2 \end{bmatrix}$$

which leads to an exceptionally conservative result: stability follows if both gain bounds are less than 1. Luckily, we can do better than this. Recall that if a system is (Q, S, R) dissipative, then it is also $(\alpha Q, \alpha S, \alpha R)$ dissipative for any constant $\alpha > 0$. With this change, we get

$$\widehat{Q} = \begin{bmatrix} -1 + \alpha k_2^2 & 0 \\ 0 & -\alpha + k_1^2 \end{bmatrix}$$

It is easy to see that an α can be found that makes $\widehat{Q}$ negative definite provided that $k_1 k_2 < 1$.

6. Passive subsystems

Let us now consider a more general interconnection of passive systems. For the single-loop case we already know that simple passivity is not good enough; we need to introduce the concepts of ISP, OSP, and VSP. Our first result shows the tradeoff between these properties.

THEOREM 16. *Suppose that $H + H^T \geq 0$, and that all subsystems are passive. Let the subsystems be ordered such that the first n_1 are VSP, the next n_2 are OSP,* *the next n_3 are ISP, and the remaining $n_4 = N - (n_1 + n_2 + n_3)$ are passive. Let H be partitioned in the obvious way as*

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \\ H_{41} & H_{42} & H_{43} & H_{44} \end{bmatrix}$$

Then a sufficient condition for stability is that the columns of

$$\begin{bmatrix} H_{13} & H_{14} \\ H_{33} & H_{34} \end{bmatrix}$$

be linearly independent.

Proof. After collecting the dissipativeness parameters of the subsystems, we have $Q = \text{diag}\{-\Lambda_1, -\Lambda_2, 0, 0\}$, $S = I$, and $R = \text{diag}\{-\Lambda_3, 0, -\Lambda_4, 0\}$, where the Λ_i are positive definite diagonal matrices. Obviously $Q \le 0$ and $H^T RH \le 0$, and we have assumed that $SH + H^T S^T \ge 0$, which means that each term in the expression for $\widehat{Q}$ is nonpositive definite. All that remains to be shown is that $\widehat{Q}$ is nonsingular.

For the first term, we have

$$y^T Qy = -y_1^T \Lambda_1 y_1 - y_2^T \Lambda_2 y_2$$

and this is negative whenever $y_1 \ne 0$ and/or $y_2 \ne 0$. That means that $y^T \widehat{Q} y = 0$ is possible, if it occurs at all, only for y of the form

$$y = \begin{bmatrix} 0 \\ 0 \\ y_3 \\ y_4 \end{bmatrix}$$

At the same time, we have

$$y^T H^T RHy = -z_1^T \Lambda_3 z_1 - z_3^T \Lambda_4 z_3$$

where $z = Hy$. For our restricted class of y vectors, we have

$$\begin{bmatrix} z_1 \\ z_3 \end{bmatrix} = \begin{bmatrix} H_{13} & H_{14} \\ H_{33} & H_{34} \end{bmatrix} \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}$$

Let

$$H_{part} = \begin{bmatrix} H_{13} & H_{14} \\ H_{33} & H_{34} \end{bmatrix}$$

Then, for the subspace of y values under consideration, we have

$$y^T \widehat{Q} y \le y^T H^T RHy = - \begin{bmatrix} y_3^T & y_4^T \end{bmatrix} H_{part}^T \begin{bmatrix} \Lambda_3 & 0 \\ 0 & \Lambda_4 \end{bmatrix} H_{part} \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}$$

and the matrix $H_{part}^T \begin{bmatrix} \Lambda_3 & 0 \\ 0 & \Lambda_4 \end{bmatrix} H_{part}$ is nonsingular if H_{part} has full column rank. This completes the proof. □

Notice that the rows of H_{part} correspond to those subsystems which are ISP or better, while the columns correspond to the subsystems that are not OSP. That gives some insight into how the different kinds of "strictly passive" are being traded off among one another.

A different approach, which does not require strict passivity, is illustrated by the following theorem.

Theorem 17. *Let all subsystems be passive, but not necessarily strongly passive; and suppose that all subsystems are single-input single-output systems. Then a sufficient condition for stability is that there exist a positive definite diagonal matrix P such that $PH + H^T P > 0$.*

PROOF. If subsystem i is passive, then it is also $(0, p_i, 0)$ dissipative for any $p_i > 0$. This gives $\widehat{Q} = -\left(PH + H^T P\right)$. □

Of course, the conditions of this theorem are unlikely to be satisfied if H is a mere interconnection matrix with ± 1 and 0 entries. The theorem only becomes interesting if there is local constant feedback around some of the subsystems.

If it were not for the condition that P be diagonal, the condition being considered here would reduce to a simple eigenvalue condition on the matrix H. The case where P has to be diagonal is somewhat different, and is covered in the Appendix. A sufficient condition for the existence of such a P turns out to be that a matrix derived from H be an M-matrix. M-matrices are also explained in the Appendix.

7. A small gain theorem

In this section we look at the case where all subsystems have finite gain.

THEOREM 18. *Let subsystem i have finite gain γ_i, for $i = 1..N$, and suppose that each subsystem has only one input and one output. Define a gain matrix $\Gamma =$ diag $\{\gamma_1, ...\gamma_N\}$, and let $A = \Gamma H$. Then the overall system is stable if there exists a diagonal positive definite matrix P such that $P - A^T P A > 0$.*

PROOF. The ith subsystem is $(-1, 0, \gamma_i^2)$ dissipative, which means that it is $(-p_i, 0, p_i\gamma_i^2)$ dissipative for any $p_i > 0$. We then get $\widehat{Q} = -P + HP\Gamma^2 H$, where $P =$diag$\{p_1, ...p_N\}$. □

Methods for checking for the existence of a suitable P are also covered in the Appendix. In particular, a sufficient — but far from necessary — condition is that the matrix $\widehat{A}$ with elements

$$\begin{aligned}\hat{a}_{ii} &= 1 - |a_{ii}| \\ \hat{a}_{ij} &= -|a_{ij}| \text{ for } j \neq i\end{aligned}$$

have positive principal minors. That is, it is an M-matrix (see Appendix). This is precisely the condition given in [**PM74**] and [**Coo74**] for input-output stability.

Variants of this result have been produced by a number of different researchers. See, for example, [**Š72**], [**AK72**], [**MP72**], [**Mic74**], [**RM76**], [**Š76**]. Some of those results turn out to be special cases of the above theorem, but others — especially those that rely on special assumptions on the state-space model — neither imply nor are implied by our present result. It would be fair to say that the results of this book are basically input-output results (even though they also imply asymptotic stability), so they do not cover cases that require special properties of a state-space model.

It is worth noting that many of the known "small gain" stability criteria reduce to an M-matrix test. (See Appendix.) The great virtue of an M-matrix test is that it is easy to check. The disadvantage is that such tests tend to be tests for *weakly coupled* subsystems. In effect, they are tests for the condition "we know that each subsystem is stable, and the interconnection does not disturb this property". This is appropriate for some practical examples. For others, we would be better off looking for conditions of the form "the feedback improves the stability".

8. Conic subsystems

If a single-input single-output system satisfies the condition

$$\langle y - au, bu - y\rangle_T \geq 0$$

for some scalars a and $b \geq a$ (and all u), then we say that it is interior conic, or conic inside the sector $[a, b]$. Similarly, if

$$\langle y - au, bu - y \rangle_T \leq 0$$

then we say that it is exterior conic, or conic outside the sector $[a, b]$. Notice that finite gain systems are special cases of interior conic systems. Moreover, as we will see in Chapters 8 and 9, conicity is often a property that can be easily checked.

Suppose now that a number of conic systems are interconnected. If the ith subsystem is inside or outside the sector $[a, b]$, then it is $(-\sigma_i, \frac{1}{2}(a_i + b_i)\sigma_i, -a_i b_i \sigma_i)$ dissipative, where $\sigma_i = +1$ for internal and -1 for external conicity. Define matrices $A = \text{diag}\{a_1, ..., a_N\}$, $B = \text{diag}\{b_1, ..., b_N\}$, and $\Sigma = \text{diag}\{\sigma_1, ..., \sigma_N\}$. Finally, let $C = \frac{1}{2}(A + B)$, and $D = \frac{1}{2}(B - A)$.

THEOREM 19. *The interconnection of conic subsystems is stable if there exists a positive definite diagonal P such that*

$$(I + CH)^T P\Sigma(I + CH) - (DH)^T P\Sigma(DH) > 0$$

PROOF. After using positive weighting factors p_i as in the earlier theorems, we have $Q = -P\Sigma$, $S = \frac{1}{2}(A + B)P\Sigma$, and $R = -ABP\Sigma$. Putting these formulae together, we get

$$\widehat{Q} = -(I + CH)^T P\Sigma(I + CH) + (DH)^T P\Sigma(DH)$$

□

When the methods of the Appendix are used to check for a suitable P, this theorem often gives the same results as the method of Porter and Michel [**PM74**]. Nevertheless, there are certainly examples where this theorem gives less conservative stability criteria than the technique in [**PM74**].

9. Examples

Suppose that we have two finite gain subsystems, each with gain $\leq \frac{1}{2}$, and an interconnection matrix

$$H = \begin{bmatrix} 1 & -1 \\ -1 & -k \end{bmatrix}$$

What values of k will preserve asymptotic stability?

From Theorem 18, a sufficient condition for stability is that there exist a diagonal $P > 0$ such that $P - A^T PA > 0$, where

$$A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & -k \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2}k \end{bmatrix}$$

Using the simplest method in the Appendix, stability follows if the matrix

$$\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 - \frac{1}{2}|k| \end{bmatrix}$$

has positive principal minors; that is, if $|k| < 1$. This is the result obtained by more conventional approaches. Alternatively, stability follows if there is a diagonal $P > 0$ such that $PF + F^T P > 0$, where $F = (I - A)(I + A)^{-1}$. Since F is a 2×2 matrix, it is easily shown that this is equivalent to requiring F to have positive principal minors. This criterion leads to the condition $-1 < k < 5/3$. This example shows that it is not always a good idea to look for the most obvious condition.

For another example, consider the system in Figure 2. This is a concatenation of three subsystems, with local feedback gains α, β, γ, plus an overall positive

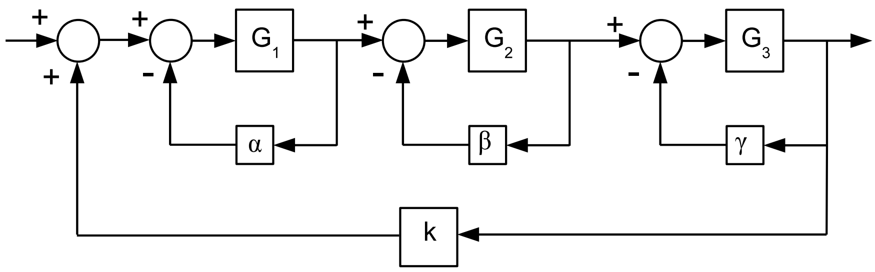

FIGURE 2. An example with three subsystems

feedback gain k. The end result is an interconnection matrix

$$H = \begin{bmatrix} \alpha & 0 & -k \\ -1 & \beta & 0 \\ 0 & -1 & \gamma \end{bmatrix}$$

Theorem 16 does not provide stability in this case, since none of the subsystems is VSP or ISP. Trying another approach, Theorem 17 predicts stability if there exists a diagonal $P > 0$ such that $PH + H^T P > 0$. A sufficient condition for this is that H be quasidominant. This is satisfied if $\alpha > 0$, $\beta > 0$, $\gamma > 0$, and $|k| < \alpha\beta\gamma$. We can, however, do better. Applying decision theory methods [**ABJ75**], [**Jac74**], we find that the necessary and sufficient condition for the existence of a suitable P is $\alpha > 0$, $\beta > 0$, $\gamma > 0$, and $-8\alpha\beta\gamma < k < \alpha\beta\gamma$.

CHAPTER 7

Instability

1. Criteria for instability

So far we have looked at a number of sufficient conditions for stability. Can we get conditions that are both necessary and sufficient? In general, no. We can, nevertheless, partly fill the gap by finding conditions that are sufficient for instability. If nothing else, this can give us some insight into the gaps between the stability and the instability results. More importantly, for our present purposes, the instability results of this chapter will help to clarify the relationships between the various kinds of dissipativeness.

The material in this chapter is based on the results in [**HM78**], [**MH79**], [**HM80c**], and [**HM83**], although of course those papers themselves cite earlier work by others on instability.

Let us recall some important definitions from Chapter 2. To define dissipativeness, we need first to define an input signal space U and an output signal space Y which are inner product spaces. This, however, is not sufficient, because for example when talking of instability we need to allow the possibility of signals with unbounded norm. To handle that problem, we define two extended spaces

$$\begin{aligned} U_e &= \{u : P_T u \in U \text{ for all } T\} \\ Y_e &= \{y : P_T y \in Y \text{ for all } T\} \end{aligned}$$

and then say that our system G maps U_e to Y_e. The smaller spaces U and Y are often called the spaces of small signals. Recall that P_T is the causal truncation operator, usually defined to be the projection that truncates a signal at time T.

To be strictly correct, we should define two different causal truncation operators P_T^u and P_T^y, because they operate on two different signal spaces; but we choose to omit the superscript because it is always obvious from context which signal the operator is being applied to.

With these definitions, we say that G is input-output stable if $u \in U$ implies $Gu \in Y$, and input-output unstable if there is at least one input that violates this condition. If we define the set

$$K(G) \triangleq \{u \in U : Gu \in Y\}$$

then G is input-output stable iff $K(G) = U$.

Let us also recall the following definition, again from Chapter 2.

Definition 17. *Let Q, S, and R be memoryless linear operators, with Q and R self-adjoint. Then the system defined by the (linear or nonlinear) operator equation $y = Gu$ is (Q, S, R) ultimately virtually dissipative (UVD) iff*

$$\langle y, Qy \rangle + 2 \langle y, Su \rangle + \langle u, Ru \rangle \geq 0$$

for all $u \in K(G)$.

It is useful to add one more definition, which has not been needed until now.

Definition 18. *If system G is (Q, S, R) ultimately virtually dissipative, and if in addition $K(G) = U$, then we say that G is (Q, S, R) ultimately dissipative.*

In other words, an ultimately dissipative system is one which is both ultimately virtually dissipative and input-output stable. Does this imply that it is also (Q, S, R) dissipative? This is what we shall discover in the next section.

As one final preliminary point, recall that G is called *causal* if $P_T^y G P_T^u u = P_T^y G u$ for all u, or more compactly if $P_T G P_T = P_T G$. Up to this point we have not needed to assume causality, except in state-space models where it is part of the definition of the state; but it will matter when talking about input-output instability.

2. The basic input-output instability results

To begin with, let us see when ultimate dissipativeness implies dissipativeness.

THEOREM 20. *Suppose that the system G is causal and (Q, S, R) ultimately dissipative, where $Q \leq 0$. Then G is (Q, S, R) dissipative.*

PROOF. Because $K(G) = U$, the condition

$$\langle y, Qy \rangle + 2 \langle y, Su \rangle + \langle u, Ru \rangle \geq 0$$

holds for all $u \in U$. Now consider an arbitrary $u \in U_e$, and for an arbitrary T define $y = Gu$, $u_1 = P_T u$, and $y_1 = Gu_1$. Then we have

$$\langle y_1, Qy_1 \rangle + 2 \langle y_1, Su_1 \rangle + \langle u_1, Ru_1 \rangle \geq 0$$

Causality tells us that $P_T y_1 = P_T G u_1 = P_T G P_T u = P_T G u = P_T y$. That is, y and y_1 have the same past history, although we cannot rule out their being different in the future. From this, and the fact that u_1 is a truncated signal, we deduce that $\langle y_1, Su_1 \rangle = \langle y, Su \rangle_T$ and $\langle u_1, Ru_1 \rangle = \langle u, Ru \rangle_T$.

For the remaining term, the fact that $Q \leq 0$ implies that Q can be factored as $Q = -M^* M$ where M is a memoryless operator and M^* is its adjoint. Then

$$\langle y_1, Qy_1 \rangle = -\langle My_1, My_1 \rangle \leq -\langle P_T M y_1, P_T M y_1 \rangle$$

Now, the fact that M is memoryless means that it is causal, and therefore $P_T M y_1 = P_T M P_T y_1 = P_T M P_T y = P_T M y$. We therefore deduce that

$$\begin{aligned} \langle y_1, Qy_1 \rangle &\leq -\langle P_T M y_1, P_T M y_1 \rangle \\ &= -\langle P_T M y, P_T M y \rangle = -\langle My, My \rangle_T = \langle y, Qy \rangle_T \end{aligned}$$

Putting these details together, we conclude that

$$\langle y, Qy \rangle_T + 2 \langle y, Su \rangle_T + \langle u, Ru \rangle_T \geq 0$$

and therefore G is (Q, S, R) dissipative. □

There does not appear to be any way to extend this result to the case of arbitrary Q, but that does not matter; we only need the case $Q \leq 0$ for our instability results.

The basic instability theorem can now be stated.

THEOREM 21. *Suppose that the system G is causal and (Q, S, R) ultimately virtually dissipative but not (Q, S, R) dissipative, where $Q \leq 0$. Then G is input-output unstable.*

PROOF. If G were input-output stable then we would have $K(G) = U$, meaning that G is (Q, S, R) ultimately dissipative. Theorem 20 then gives us a contradiction. □

In the case $Q < 0$, and assuming causality, we have a pleasing symmetry: dissipativeness implies stability, but ultimate virtual dissipativeness, without dissipativeness, implies instability. In the sign-indefinite case, UVD without dissipativeness implies instability, but the best we can do for a stability result is to say that dissipativeness sometimes implies stability.

3. A state-space instability result

Input-output stability is all about the relationship between inputs and outputs. If a state-space model is available, the results are implicitly for the case of zero initial state. (Or, more generally, for an initial state that is an equilibrium state.) It is, of course, possible to develop a parametrized model where $G(x_0)$ represents the input-output map when the initial state is x_0. For models of that type, the interested reader might want to look at [**HM80b**] or [**MCS82**]. The results tend to be of the form "if $G(0)$ is dissipative, then $G(x_0)$ is weakly dissipative for all reachable x_0". If you are sufficiently diligent, you will discover that you can prove "weak" versions of most of the results in this book. We have chosen to omit those results, on the grounds that a list of all the variants becomes tedious to read, and not especially interesting.

When working with a state-space model, most people prefer to ignore the input-output map and look instead at the state trajectory with zero input. There are at least three possibilities:

(1) With zero input, and any initial state, the state $x(t)$ remains bounded for all t. We then say that the system is stable in the sense of Lyapunov.
(2) With zero input, and any initial state, the state $x(t)$ remains bounded for all t, and tends to zero as $t \to \infty$. We then say that the system is asymptotically stable in the sense of Lyapunov.
(3) With zero input, there are some initial states $x(0)$ for which $\|x(t)\|$ grows without bound. In that case, the system is unstable in the sense of Lyapunov.

We can expand this list if, for example, we want to assert that the convergence or the growth is exponential.

When considering the differences between the definitions of input-output stability and Lyapunov stability, one might be forgiven for thinking that they have little to do with each other. We have already seen, however, that the input-output stability and the state-space stability results for dissipative systems are very closely related. In particular, if a system is finite-gain stable, then with a few extra well-posedness conditions it is also asymptotically stable in the sense of Lyapunov [**HM80a**]. This has long been known for linear systems, so it should not be too surprising that it also works for nonlinear systems.

The approximate state-space equivalent of the ultimate virtual dissipativeness property is cyclodissipativeness, so we should expect something along the lines of saying that if a system is cyclodissipative but not dissipative, and $Q \leq 0$, then it is unstable in some sense. We can indeed find such a result.

THEOREM 22. *Suppose that the system G is (Q, S, R) cyclodissipative but not (Q, S, R) dissipative, where $Q \leq 0$. Then G is not asymptotically stable in the sense of Lyapunov.*

PROOF. Cyclodissipativeness implies the existence of a virtual storage function $\phi(x)$ such that

$$\phi(x(t_0)) + \int_{t_0}^{t_1} \left(y^T Q y + 2y^T S u + u^T R u\right) dt \geq \phi(x(t_1))$$

and when $u(t) = 0$ for $t \geq t_0$ this reduces to

$$\phi(x(t_0)) + \int_{t_0}^{t_1} y^T Q y \, dt \geq \phi(x(t_1))$$

If $Q \leq 0$, this implies that $\phi(x(t))$ is monotonically non-increasing. If the origin is asymptotically stable, then $\lim_{t\to\infty} x(t) = 0$, so $\lim_{t\to\infty} \phi(x(t)) = 0$, therefore

$\phi(x(t_0)) \geq 0$. Since this argument holds for arbitrary $x(t_0)$, we deduce that $\phi(x) \geq 0$ for all x. In other words, this virtual storage function is actually a storage function, so that the system is (Q, S, R) dissipative. This is a contradiction, therefore our assumption of asymptotic stability is wrong. □

Note that the theorem statement does not explicitly require causality. This is because state-space models are automatically causal.

To get some intuitive feel of what this means, consider the case of a linear system. In a later chapter it will be shown that dissipativeness can be described in the frequency domain in terms of the matrix

$$M(s) = G(s)^* QG(s) + G(s)^* S + SG(s) + R$$

where $G(s)$ is the transfer function of the system being investigated. It will be shown that dissipativeness is equivalent to $M(s) \geq 0$ for all s in Re $s \geq 0$, and that cyclodissipativeness is equivalent to $M(j\omega) \geq 0$ for all real ω. (Obviously the latter condition is easier to check.) That is, dissipativeness requires checking $M(s)$ in the entire right half-plane, while cyclodissipativeness requires checking only the left boundary of that half-plane. There are well-known results in complex algebra that connect the value of a function in a region with the values on the boundary of that region, and those results depend on the number of poles inside that region.

Informally, we can say that cyclodissipativeness and dissipativeness both require checking the frequency response ($G(j\omega)$ for real ω), and that the difference between them depends on the poles of a function in the right half of the complex plane. From that point of view, it is hardly surprising that the condition "cyclodissipative but not dissipative" is related to the existence of unstable modes.

Of course, these conclusions work only in the case $Q \leq 0$. There are no corresponding results for more general Q. Nevertheless, the case of more general Q becomes relevant when we start looking at interconnected systems.

4. Interconnected systems

For an interconnected system, we use the same preliminaries as were used to obtain the stability results. We have a collection of subsystems $\{G_i, i = 1..N\}$, where G_i is (Q_i, S_i, R_i) cyclodissipative or ultimately virtually dissipative, depending on whether we are talking about an input-output or a state-space model. (To avoid tortuous paraphrases, we will use "cyclodissipative" for both cases for now.) The individual subsystems are described by $y_i = G_i u_i$. We collect those subsystem inputs and outputs into large vectors

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

and combine the cyclodissipativeness parameters into block diagonal matrices

$$\begin{aligned} Q &= \text{diag}\{Q_1, ..., Q_N\} \\ S &= \text{diag}\{S_1, ..., S_N\} \\ R &= \text{diag}\{R_1, ..., R_N\} \end{aligned}$$

Then obviously the overall system is (Q, S, R) cyclodissipative. Next, we define an interconnection

$$u = u_{ext} - Hy$$

where u_{ext} represents the external inputs, and H is a large matrix which (in most but not all applications) has mostly 0 and ± 1 entries. Now the UVD condition is

$$\langle y, Qy \rangle + 2 \langle y, Su \rangle + \langle u, Ru \rangle \geq 0$$

for the input-output scenario, and a similar condition involving virtual storage functions for the state-space scenario. Substituting the equation for u, this becomes

$$\begin{aligned} &\left\langle y, \left(Q - SH - H^T S^T + H^T RH\right) y \right\rangle \\ &\quad + 2 \left\langle y, \left(S - H^T R^T\right) u_{ext} \right\rangle + \langle u_{ext}, Ru_{ext} \rangle \geq 0 \end{aligned}$$

or

$$\left\langle y, \hat{Q} y \right\rangle + 2 \left\langle y, \hat{S} u_{ext} \right\rangle + \left\langle u_{ext}, \hat{R} u_{ext} \right\rangle \geq 0$$

where

$$\begin{aligned} \hat{Q} &= Q - SH - H^T S^T + H^T RH \\ \hat{S} &= S - H^T R^T \\ \hat{R} &= R \end{aligned}$$

Clearly the overall system, with input u_{ext} and output y, is $(\hat{Q}, \hat{S}, \hat{R})$ ultimately virtually dissipative, or cyclodissipative, depending on whether we are looking at the input-output model or the state space model. But is it also $(\hat{Q}, \hat{S}, \hat{R})$ dissipative? This is where a more careful analysis is needed.

For our next result, we need to perform the thought experiment of choosing an external input u_{ext} that will force the individual inputs u_i to desired values. In principle this is easy, because we have as many equations as unknowns. If we want $u_i = \bar{u}_i$ for all i, we just have to set

$$u_{exti} = \bar{u}_i + \sum_{j=1}^{N} H_{ij} \bar{y}_j$$

for all i, where $\bar{y}_j = G_j \bar{u}_i$. Once we have chosen these external inputs, the solution for the u_i is obtained by solving the set of equations

$$\begin{aligned} u_i &= u_{exti} - \sum_{j=1}^{N} H_{ij} y_j = \bar{u}_i + \sum_{j=1}^{N} H_{ij} \left(\bar{y}_j - y_j\right) \\ y_i &= G_i u_i \end{aligned}$$

Obviously one solution is given by $u_i = \bar{u}_i$ (and therefore $y_i = \bar{y}_i$) for all i, but is it the only solution? When all subsystems are linear, the question of uniqueness reduces down to checking that $I + HG(s)$ is invertible. For nonlinear systems, there is no such simple answer.

Luckily, a little thought shows that we do not actually need uniqueness. If the system equations have multiple solutions, we can call the overall system stable only if *all* of the multiple solutions are stable solutions. Conversely, to prove instability we need only show that any one of the multiple solutions exhibits unstable behaviour. That is, we can choose the solution that interests us, and ignore the others.

With these preliminaries out of the way, we have the following result.

THEOREM 23. *Suppose that the overall interconnected system is causal, and that subsystem G_i is (Q_i, S_i, R_i) ultimately virtually dissipative, for $i = 1..N$, but at least one subsystem G_k is not (Q_k, S_k, R_k) dissipative. Then the overall system is input-output unstable if $\hat{Q}$ as defined above satisfies $\hat{Q} \leq 0$, and one or both of the following conditions holds:*

(1) *at least one of the subsystems that is not dissipative is linear;*

(2) *for each subsystem, except possibly one of the non-dissipative ones, either* G_i *is unbiased in the sense* $G_i 0 = 0$*, and/or* $Q_i \leq 0$.

PROOF. By construction, we have

$$\left\langle y, \hat{Q}y \right\rangle_T + 2\left\langle y, \hat{S}u_{ext} \right\rangle_T + \left\langle u_{ext}, \hat{R}u_{ext} \right\rangle_T$$
$$= \sum_{i=1}^{N} \langle y_i, Q_i y_i \rangle_T + 2\langle y_i, S_i u_i \rangle_T + \langle u_i, R_i u_i \rangle_T$$

but for subsystem G_k we also know that there is some T and some $\bar{u}_k$ and some $\bar{y}_k = G\bar{u}_k$ such that

$$\langle \bar{y}_k, Q_k \bar{y}_k \rangle_T + 2\langle \bar{y}_k, S_k \bar{u}_k \rangle_T + \langle \bar{u}_k, R_k \bar{u}_k \rangle_T < 0$$

Now choose the external input u_{ext} such that $u_k = \bar{u}_k$, and $u_i = 0$ for all $i \neq k$. From our earlier remarks, it does not matter if there are multiple solutions for this choice of u_{ext}; all that matters is that there is one solution that agrees with our specification.

If G_k is linear, then we can scale up the inputs to make $u_k = \lambda \bar{u}_k$ for some scalar λ. By linearity, y_k will be scaled by the same factor, and the corresponding energy term will be scaled by λ^2. Obviously, we can make λ large enough so that the negative contribution from G_k will dominate the above sum, to produce a net negative result. It follows that the overall system cannot be $(\hat{Q}, \hat{S}, \hat{R})$ dissipative.

If G_k is not linear, then the alternative assumption of the theorem implies that $\langle y_i, Q_i y_i \rangle_T \leq 0$ for all $i \neq m$, because for each i we have either $y_i = 0$ (the unbiased case) or $Q_i \leq 0$. Again, the sum is negative and the overall system cannot be $(\hat{Q}, \hat{S}, \hat{R})$ dissipative.

In either case, then, we conclude instability from Theorem 21. □

Our next task, obviously, is to find a comparable result for state-space instability. Here is one possible way of doing it.

THEOREM 24. *Suppose that the overall interconnected system has a state-space representation, that subsystem* G_i *is* (Q_i, S_i, R_i) *cyclodissipative, for* $i = 1..N$*, that at least one subsystem* G_k *is not* (Q_k, S_k, R_k) *dissipative, and that* $\hat{Q}$ *as defined above satisfies* $\hat{Q} \leq 0$*. Then the overall system is not asymptotically stable in the sense of Lyapunov.*

PROOF. Obviously the overall system is $(\hat{Q}, \hat{S}, \hat{R})$ cyclodissipative. With zero external input, we have

$$\phi(x(t_0)) + \int_{t_0}^{t_1} y(t)^T \hat{Q} y(t) dt \geq \phi(x(t_1))$$

whenever $t_1 \geq t_0$. With $\hat{Q} \leq 0$, that implies that $\phi(x(t))$ is monotonically non-increasing. Because one of the subsystems is cyclodissipative but not dissipative, there exists at least one x such that $\phi(x) < 0$. If we choose that x as the initial state, then the state can never converge to a state where $\phi(x) = 0$. □

Observe that Theorem 24 does not rule out the possibility that $\phi(x(t))$ converges to a finite constant. This could mean that the state trajectory converges to a limit cycle. It could also mean that the state converges to a stable equilibrium, but that stable equilibrium is not the origin. It is not uncommon for a nonlinear system to have multiple equilibria, and this includes the case where the origin is locally unstable.

Stronger results are possible if we know more about the properties of the virtual storage functions. For example, if the non-dissipative subsystem is linear, then it has a virtual storage function that is quadratic. That means that there are states arbitrarily close to the origin where $\phi(x) < 0$. It also means that there is a cone of states in which $\lim_{x\to\infty} \phi(x) = \infty$.

Stronger results are also possible if we strengthen the matrix condition to $\hat{Q} < 0$. In that case, however, we need to impose an observability-like condition to cover the complications that can occur when $y(t) \to 0$.

CHAPTER 8

Frequency domain tests

1. Introduction

Although this book is primarily about nonlinear systems, it must be remembered that nonlinear means “not necessarily linear”. Linear systems form an important special subcase. It is a remarkable fact that, when the general nonlinear systems results are specialised to the linear case, one sometimes gets better results than have been obtained in the literature by assuming linearity from the outset. One may conjecture that, for at least some problem areas, linearity has been a red herring which has led researchers into convoluted algebra that served to obscure some key issues.

Linear systems have been mentioned in previous chapters in the context of algebraic tests for dissipativeness. In this chapter we concentrate on frequency domain tests: criteria for dissipativeness that look at the the transfer function or transfer function matrix of a linear system. Transfer functions are, of course, input-output descriptions, but this is no real restriction; there are well-known equivalences between the internal and external properties of linear systems, provided only that a minimal state-space representation is chosen. Even in the non-minimal case, there are equally well-known methods for separating out the uncontrollable and/or unobservable parts of a linear system and treating them separately.

The key virtue of a transfer function description is that it lends itself easily to graphical interpretations. Given a transfer function $G(s)$, one can extract a great deal of information from the plot of $G(j\omega)$ as ω varies from $-\infty$ to $+\infty$. This leads to simple hand calculations in the case of low-order systems, or — for more complex systems — to tests which are easily automated with the aid of suitable graphics software. For historical reasons, transfer function tests also fit in well with the intuitive mental processes of a control system designer. If, for example, a graphical stability test fails, a designer can often judge by looking at the graph just how the system needs to be modified.

Most of the results of this chapter fall into the class of what are known as “circle criteria”. Historically, one tends to associate this name with the stability tests of Sandberg [**San64**] and Zames [**Zam66**]. The general flavour of the results is as follows. Given a feedback system in which the forward path is linear with scalar transfer function $G(s)$, and the feedback path is a sector nonlinearity, stability can be deduced if the graph of $G(j\omega)$ avoids a circle in the complex plane, and encircles it the correct number of times. The location and size of the critical circle depend on the sector bounds of the nonlinearity. In some cases, the circle degenerates to a straight line. In the special case where the feedback is a constant linear gain, the circle shrinks to a point, and the test reduces to the classical Nyquist criterion. Thus, this sort of circle criterion can be thought of as a generalisation of the Nyquist criterion. Interested readers may also care to look at the relationship of the circle criterion to Aizerman’s conjecture.

If there is more than one nonlinearity, it is always possible to collect all the nonlinearities together and call the collection a single multivariable nonlinearity;

and to re-draw the block diagram of the system in the form of a single-loop feedback system. This, however, makes $G(s)$ a matrix. Similar stability criteria still apply to this case, but it is less obvious how to turn them into graphical tests.

One approach to this problem is provided by the multivariable circle criterion of Rosenbrock [**Ros70**]. The general idea is to plot the diagonal elements of $G(j\omega)$; but the plots are not simple curves in the complex plane, they are bands whose width depends on the off-diagonal elements. Several variants of this idea have been developed, of which the best known is probably that of Cook [**Coo74**].

A problem with multivariable circle criteria is that they require *diagonal dominance*: the off-diagonal elements of $G(s)$ must be in some sense small with respect to the diagonal elements. If one is free to design pre-compensators which modify $G(s)$ in order to obtain the required properties, then this might not be a problem. There are, however, cases where this freedom is not available. Notice, too, that if a system with multiple unrelated nonlinearities is re-drawn in the form of a feedback system, then the relationship of the resulting $G(s)$ — which arises from lumping together all the linear subsystems — to the original system model is not always a simple one, so there is no good reason to expect it to have a diagonal dominance property.

Our solution to this problem is to make a subtle shift in philosophical emphasis. Historically, circle criteria have been seen as tests for *stability*. In this chapter, we consider circle criteria to be tests for *dissipativeness*. The implications of this distinction might not be immediately obvious, since the known circle criteria were in any case derived using concepts very much akin to dissipativeness. The key advantage of the shift in viewpoint is that it frees us from having to formulate the stability problem in terms of a single-loop system. Given a complicated system of many interconnected parts, one can break it into subsystems in any way in which convenience dictates. Often there will be an "obvious" decomposition suggested by the physical nature of the system. Dissipativeness tests — which are circle criteria in the case of linear systems — can be applied to each subsystem independently, without having to look at how that subsystem is related to the others. Then, if stability is the property of interest, once can apply the tests given in an earlier chapter, which relate stability of an interconnected system to the dissipativeness parameters of its subsystems.

This is not to say that multivariable circle criteria are unimportant. A "good" decomposition of a system can still produce subsystems with multiple inputs and outputs; therefore we must still consider the multivariable case. We are, however, freed from the artificial aggregation of subsystems with no purpose other than to reformulate the problem as a single-loop problem.

Actually, most of the graphical tests in this chapter are for cyclodissipativeness rather than for dissipativeness. (To be more technically correct, the tests are for the input-output property of ultimate virtual dissipativeness; but, for linear systems with a minimal state-space representation, cyclodissipativeness and ultimate virtual dissipativeness are equivalent.) Dissipativeness tests follow as corollaries of the cyclodissipativeness tests. The reason for this is that, broadly speaking, dissipativeness depends on the behaviour of $G(s)$ in the entire half-plane Re $s \geq 0$, whereas cyclodissipativeness depends only on the behaviour of $G(j\omega)$ for real ω. Since $G(j\omega)$ is easily displayed in graphical form, as a function of the real variable ω, tests for dissipativeness are most easily done by checking for cyclodissipativeness as an intermediate step. As was seen in Chapter 7, tests for cyclodissipativeness are also of interest in their own right, in connection with instability theorems.

2. General frequency domain criteria

The systems of interest in this chapter have state-space representation

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{7}$$

where $u \in L_{2e}^m$, $y \in L_{2e}^p$, and $x \in R^n$. We refer to this system as system G. Its transfer function is, of course,

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

In a later section, the discrete-time counterpart of system 7 will be considered.

A standing assumption throughout this chapter is that the state-space representation is minimal. We shall be looking at tests based on the transfer function $G(s)$, and such tests are inherently incapable of giving information about uncontrollable or unobservable modes. In case such modes exists, the easiest way to deal with them is to decompose the state equation into several interconnected subsystems, and to use frequency domain methods for the controllable and observable subsystem.

Given that we are interested in the integral

$$\int_0^T w(u(t), y(t))dt = \int_0^T \left(y(t)^T Q y(t) + 2y(t)^T S u(t) + u(t)^T R u(t)\right) dt$$

it is natural to look at the frequency domain quantity

$$Y(s)^* Q Y(s) + 2Y(s)^* S U(s) + U(s)^* R U(s) = U(s)^* M(s) U(s)$$

where as usual $U(s)$ and $Y(s)$ denote Laplace transforms, and the star means adjoint (=complex conjugate transpose, in the case of complex matrices). The matrix $M(s)$ has the form

$$M(s) = R + S^T G(s) + G(s)^* S + G(s)^* Q G(s)$$

Our aim is to establish a relationship between dissipativeness and the condition $M(s) \geq 0$ for Re $s \geq 0$. For cyclodissipativeness, the weaker condition $M(j\omega) \geq 0$ for $\omega \in R$ will be sufficient. Informally, this is because cyclodissipativeness can be thought of as a constraint on the periodic motions of the system; that is, on the steady-state response to a periodic input. For dissipativeness, we need to consider the total response including transients; furthermore, these "transients" can be exponentially unbounded in the case of an unstable system.

Observe that, for any complex s, $M(s)$ is self-adjoint. The notation $M(s) \geq 0$ means $y^* M(s) y \geq 0$ for all complex y.

A formal proof for the dissipativeness result requires some care, essentially because signals in L_{2e} are not necessarily Laplace transformable. The cyclodissipativeness result is easier, and can be stated immediately.

THEOREM 25. *System G is (Q, S, R) cyclodissipative if and only if $M(j\omega) \geq 0$ for all real ω such that $j\omega$ is not a pole of $G(s)$.*

PROOF. First, suppose that the system is cyclodissipative. Choose any ω_0 such that $j\omega_0$ is not a pole of G, and let $u(t) = \text{Re } ke^{j\omega_0 t}$, for an arbitrary complex k of appropriate dimension. If we choose $x(0) = \text{Re } (j\omega_0 I - A)^{-1}Bk$, then it is easy to see that $x(t)$ is periodic and that $y(t) = \text{Re } G(j\omega_0)ke^{j\omega_0 t}$. Notice that $x(0) \neq 0$ in general, but this does not matter; the important point is that $x(T) = x(0)$, where T is the period of the input.

Now, the evaluation of

$$\int_0^T w(u(t), y(t))dt = \int_0^T \begin{bmatrix} y(t)^T & u(t)^T \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} dt$$

is a simple calculation involving sines and cosines. The average value of the integrand turns out to be $\frac{1}{2}k^* M(j\omega_0)k$. Cyclodissipativeness implies that the integral is nonnegative over any whole number of periods, thus this average value must be nonnegative. Since k was arbitrary, we deduce that $M(j\omega_0) \geq 0$ for arbitrary ω_0.

For the converse, we can use Parseval's theorem, which states that

$$\int_0^\infty f(t)^T g(t)dt = \frac{1}{2\pi}\int_{-\infty}^\infty F(j\omega)^* G(j\omega)d\omega$$

for any $f, g \in L_2^n$, where as usual the upper-case quantities are Laplace transforms. We can apply this with

$$f(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} \qquad g(t) = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\begin{bmatrix} y(t) \\ u(t) \end{bmatrix}$$

Then clearly the frequency domain integral is nonnegative if $M(j\omega) \geq 0$ for all ω; but of course this result only applies if u is chosen such that both u and y are square integrable.

Let $T \geq 0$ be arbitrary, choose $x(0) = 0$, and choose any $u \in L_{2e}^m[0, T]$ such that $x(T) = 0$. We can extend the definition of u to make $u(t) = 0$ for all $t > T$. Clearly these signals are such that Parseval's theorem may be applied. The conclusion is that nonnegativity of $M(j\omega)$ implies

$$\int_0^T w(u(t), y(t))dt \geq 0$$

for the class of $u \in L_{2e}^m[0, T]$ for which $x(T) = x(0) = 0$. This concludes the proof. □

In the sequel, we shall have frequent recourse to the following assumption.

Assumption 2. *The supply rate $w(u, y)$ is such that, for any $y \neq 0$, there exists a choice of $u = k(y)$ such that $w(k(y), y) < 0$.*

To see the point of this assumption, consider an extreme case of a supply rate where Q, S, and R are chosen in a way that gives $w(u, y) \geq 0$ for all u and y. This would lead to

$$\int_0^T w(u(t), y(t))dt \geq 0$$

for all u, all $T \geq 0$, and all $x(0)$, independently of any property of the system under consideration. Dissipativeness with respect to such a supply rate is not an interesting property, since *every* system is dissipative with respect to that w. At the other extreme, a condition like $w(u, y) \leq 0$ for all u and y would mean that only very trivial systems could be dissipative. Assumption 2 steers a course between these two extremes. In effect, the assumption defines the class of "interesting" supply rates.

The quadratic nature of $w(u, y)$ means that, if there is any u such that $w(u, y) < 0$, then there is such a u which has the form $u = -Ky$. Thus, an alternative formulation of the assumption is to demand the existence of a constant matrix K such that

$$Q - SK - K^T S^T + K^T RK < 0 \tag{8}$$

In the following chapter, we shall see that this reduces down to the condition $S^2 - QR > 0$ for single-input single-output systems, in which case the matrices are 1×1 matrices. There does not appear to be a corresponding simplification in the general matrix case.

Assumption 2 need not be a blanket assumption; it will be invoked only as needed. In this chapter, the assumption is needed only to establish a relationship between dissipativeness and the behaviour of $M(s)$ for s in the right half-plane.

The next result provides a link between cyclodissipativeness and dissipativeness for linear systems. For convenience, it is developed in two steps: first for the case where $Q < 0$, and then for the general case.

LEMMA 5. *Suppose that G is observable and (Q, S, R) cyclodissipative where $Q < 0$. Then G is (Q, S, R) dissipative iff the state-space representation of G is asymptotically stable.*

PROOF. From the results of chapter 3, cyclodissipativeness implies the existence of $\phi(x) = x^T P x$ such that

$$\phi(x(t_0)) + \int_{t_0}^{t_1} w(u(t), y(t)dt \geq \phi(x(t_1))$$

With $u(t) = 0$ for all $t \geq t_0$, the condition $Q < 0$ implies $w(0, y) \leq 0$, and asymptotic stability implies that $\phi(x(t_1)) \to 0$ as $t_1 \to \infty$. Taking the limit as $t_1 \to \infty$, it follows that $\phi(x(t_0)) \geq 0$ for all $x(t_0)$, which means that G must be dissipative with respect to the given supply rate.

The converse result is easy: if G is dissipative, then $\phi(x)$ is a Lyapunov function which establishes asymptotic stability. □

It would appear, then, that the difference between dissipativeness and cyclodissipativeness lies in the stability properties of G. This statement is partially true; and in fact Lemma 5 lies at the heart of almost all the stability and instability results in this book. It must be noted, however, that the lemma applies only to the case $Q < 0$. More generally, it is quite possible for an unstable system to be dissipative.

The following theorem gives a more precise characterisation of dissipativeness. For the supply rates of interest to us, linear dissipative systems are precisely those linear systems which can be stabilised by constant linear output feedback.

THEOREM 26. *Suppose that G is (Q, S, R) cyclodissipative, and let K be a constant matrix chosen such that $(I + DK)$ is nonsingular and $Q - SK - K^T S^T + K^T RK < 0$. Then G is (Q, S, R) dissipative if and only if the system with feedback $u = -Ky$ is asymptotically stable.*

This is a key result which underlies the dissipativeness tests in the following sections. Briefly, the strategy is to use graphical tests for cyclodissipativeness, and then to use Theorem 26 to turn them into tests for dissipativeness. Before proceeding to a proof of this result, it is appropriate to look at its assumptions. The condition on $(I + DK)$ is a minor technicality which ensures that the feedback system is well-posed. The condition is of minor importance; if $(I + DK)$ were singular, it would always be possible to perturb K slightly, to make $(I + DK)$ nonsingular without disturbing the other conditions of the theorem. The important constraint is the negative definiteness condition. It can be seen that such a K can be found precisely when the triple (Q, S, R) falls within our class of "interesting" triples. Thus, our definition of "interesting" defines the class of supply rates for which Theorem 26 is applicable.

Finally, note that the theorem works with *any* K such that $Q - SK - K^T S^T + K^T RK < 0$. There is no question of having to search among all possible feedbacks to find one that is stabilising. It turns out — indeed, it is a consequence of the theorem — that among the set of all K that satisfy this inequality, either all are stabilising or none of them are. Thus, any K in this set is as good as any other.

Proof of Theorem 26.

Let $u = u_1 - Ky$, and let G_1 denote the map from u_1 to y. We have

$$\begin{aligned} & \begin{bmatrix} y^T & u^T \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} \\ = & \begin{bmatrix} y^T & u_1^T \end{bmatrix} \begin{bmatrix} I & -K^T \\ 0 & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} I & 0 \\ -K & I \end{bmatrix} \begin{bmatrix} y \\ u_1 \end{bmatrix} \\ = & \begin{bmatrix} y^T & u_1^T \end{bmatrix} \begin{bmatrix} Q - SK - K^T S^T + K^T RK & S - K^T R \\ S^T - RK & R \end{bmatrix} \begin{bmatrix} y \\ u_1 \end{bmatrix} \end{aligned}$$

and therefore G is (Q, S, R) (cyclo)dissipative if and only iff G_1 is (Q_1, S_1, R) (cyclo)dissipative, where $Q_1 = Q - SK - K^T S^T + K^T RK$ and $S_1 = S - K^T R$. Since G_1 is in the class of systems to which Lemma 5 is applicable, the result follows immediately.

The results given so far in this section are sufficient to develop graphical criteria. For conceptual completeness, however, we need one more result. As before, it is convenient to have a separate treatment of the case $Q < 0$.

LEMMA 6. *If $Q < 0$, then G is (Q, S, R) dissipative if and only if $M(s) \geq 0$ for all s in Re $s \geq 0$.*

PROOF. Note first that, for s close to a pole of $G(s)$, the expression for $M(s)$ is dominated by the term $G(s)^* QG(s)$. If $Q < 0$, then $G(s)$ can have no pole in the region where $M(s) \geq 0$. (For a more formal proof of this fact, one can prove a finite upper bound for $\|G(s)\|$ in the region where $M(s) \geq 0$.) Thus, the conditions $Q < 0$ and $M(s) \geq 0$ in the right half-plane imply that G is asymptotically stable. Then $M(s) \geq 0$ in the region Re $s \geq 0$ implies that $M(j\omega) \geq 0$ for all real ω, which in turn implies that G is cyclodissipative, by Theorem 25. Finally, Lemma 5 tells us that G is dissipative.

For the converse, suppose that G is dissipative; then of course G is also cyclodissipative, which implies by Theorem 25 that $M(j\omega) \geq 0$ for all real ω. It also implies, by Lemma 5, that $G(s)$, and therefore $M(s)$, has no poles in the closed right half-plane. It is a standard result of complex analysis that the conditions $M(j\omega) \geq 0$ for ω real, and $M(s) \geq 0$ for Re $s \geq 0$, are equivalent given this constraint on the poles. Therefore, dissipativeness implies the desired frequency domain condition. □

For more general Q, this argument fails because $G(s)$ might have right half-plane poles. Fortunately, there is another route to the desired result.

THEOREM 27. *If there exists some K such that $Q - SK - K^T S^T + K^T RK < 0$, then G is (Q, S, R) dissipative if and only if $M(s) \geq 0$ for all s in Re $s \geq 0$ such that s is not a pole of G.*

PROOF. As in the proof of Theorem 26, let G_1 represent the system with feedback $-Ky$, and note that G is (Q, S, R) (cyclo)dissipative iff G_1 is (Q_1, S_1, R) (cyclo)dissipative, where Q_1 and S_1 were defined earlier. The transfer function of G_1 is $G_1(s) = G(s) \left[I + KG(s)\right]^{-1}$.

Consider the function

$$M_1(s) = R + S_1^T G_1(s) + G_1(s)^* S_1 + G_1(s)^* Q_1 G_1(s)$$

By expanding out the terms, we find that

$$\left[I + KG(s)\right]^* M_1(s) \left[I + KG(s)\right] = M(s)$$

and thus $M_1(s) \geq 0$ if and only if $M(s) \geq 0$, except possibly at the poles of $G(s)$ and the poles of $G_1(s)$. The poles of $G_1(s)$ may in fact be ignored in this argument, since they are confined to the left half-plane.

The proof is now immediate, via the equivalences

$$\begin{aligned} G \text{ is } (Q,S,R) \text{ dissipative} \quad &\Leftrightarrow \quad G_1 \text{ is } (Q_1,S_1,R) \text{ dissipative} \\ &\Leftrightarrow \quad M_1(s) \geq 0 \text{ in Re } s \geq 0, \text{ by Lemma 6} \\ &\Leftrightarrow \quad M(s) \geq 0 \text{ in Re } s \geq 0 \end{aligned}$$

where, in the last line, it is understood that we ignore those s that are poles of G. □

Actually, it does not matter whether the closed half-plane Re $s \geq 0$ or the open half-plane Re $s > 0$ is used in the statement of this theorem, because $M(s)$ is continuous except at the poles of $G(s)$. For the same reason, we do not have to assert anything about how $M(s)$ behaves at the poles of $G(s)$. Informally, we do not mind if $M(s)$ becomes infinite, provided that it goes to $+\infty$ rather than to $-\infty$; and this is ensured by the condition that $M(s) \geq 0$ in the vicinity of the right half-plane poles.

It is legitimate to ask whether it is really necessary to assume that there exists some K such that $Q - SK - K^T S^T + K^T RK < 0$. Perhaps surprisingly, there are cases known where the theorem fails in the absence of that assumption. That is, we really are restricted to the "interesting" supply rates in order to get a simple correspondence between time-domain and frequency-domain properties.

The formulation of equation 7 precludes the possibility of $G(s)$ having a pole at infinity. To allow that case, we would have to modify 7 to include derivatives of the input; this would lead us into a complicated redefinition of the class of admissible inputs, a digression which it is probably best to avoid. As a practical matter, though, there is a simple modification to the frequency domain criteria that handles this case. Instead of looking at $M(j\omega)$ for real ω, plot $M(s)$ for s moving clockwise along a D-shaped contour, which consists of the line $s = j\omega$ for $\omega \in [-k, k]$, and a semicircle of radius k in the right half-plane; and take the limit as $k \to \infty$.

It can also be convenient, especially in computer implementation of the graphical tests, to have the contour avoid the imaginary poles of $G(s)$, if any. To do this, let the contour follow a semicircle in the right half-plane, of radius ε, around each imaginary pole; and take the limit as $\varepsilon \to 0$. These refinements, which should be familiar to anyone who knows the Nyquist stability criterion, are necessary in the tests later in this chapter which require one to count how many times the plot of $G(j\omega)$ encircles a critical region.

3. Graphical tests: the scalar case

In this section, we focus attention on linear systems with a scalar transfer function; that is, with a single input and a single output. This case is singled out for special attention because (a) it is the case which most commonly occurs in practice; (b) the results are in part prerequisites for the multivariable results to be given later; and (c) for a scalar transfer function, it is possible to give a complete catalogue of all possible dissipativeness tests.

If $G(s)$ is a scalar function of s, then also Q, S, and R in the dissipativeness condition are scalars. It should be clear that — even in the non-scalar case — a system is (Q, S, R) (cyclo)dissipative if and only if it is $(\alpha Q, \alpha S, \alpha R)$ (cyclo)dissipative, for any real constant $\alpha > 0$. This means that it is only necessary to look at the three cases $Q > 0$, $Q < 0$, and $Q = 0$. In the case $Q = 0$, there are two subcases depending on the sign of S.

Given that all quantities are scalars, we can rewrite the inequality

$$R + S^T G(j\omega) + G(j\omega)^* S + G(j\omega)^* Q G(j\omega) \geq 0$$

in the form

$$Q\left|G(j\omega)+\frac{S}{Q}\right|^2 \geq \frac{S^2}{Q}-R$$

provided that $Q \neq 0$. If $Q > 0$, this becomes

$$\left|G(j\omega)+\frac{S}{Q}\right|^2 \geq \frac{S^2-QR}{Q^2}$$

In the case $Q < 0$, we have the same condition but with the inequality sign reversed. Finally, if $Q = 0$ we have the simpler inequality

$$2S \text{ Re } G(j\omega) \geq -R$$

whose interpretation depends on the sign of S. The special case $Q = S = 0$ can, of course, be ignored as being of no practical interest.

In summary, Theorem 25 leads to the following test for cyclodissipativeness.

THEOREM 28. *The necessary and sufficient conditions for a scalar G to be (Q, S, R) cyclodissipative are:*

(1) *If $Q > 0$, the graph of $G(j\omega)$ lies outside the circle with centre $-S/Q + j0$ and radius $\frac{1}{Q}\sqrt{S^2 - QR}$.*
(2) *If $Q < 0$, the graph of $G(j\omega)$ lies inside the circle with centre $-S/Q + j0$ and radius $\frac{1}{|Q|}\sqrt{S^2 - QR}$.*
(3) *If $Q = 0$, the graph of $G(j\omega)$ lies to the right (if $S > 0$) or to the left (if $S < 0$) of the vertical line Re $s = -\frac{R}{2S}$.*

The name "circle criterion" can be justified in case 3 if we think of a straight line as a degenerate case of a circle with infinite radius.

For a given $G(s)$, it is evident that the choice of Q, S, and R is far from being unique. It is common to find, for example, that the graph of $G(j\omega)$ lies inside one circle and outside another. In applying these tests with a view to, say, checking stability, a certain amount of judgement may be needed in selecting the "best" circle.

Notice, by the way, that the frequency domain inequalities are of a "greater than or equal" form; strict inequality is not needed. In applying Theorem 28, the graph of $G(j\omega)$ is allowed to touch the circle or straight line.

Another point to notice is that case 2, with $Q < 0$, requires $|G(j\omega)|$ to be bounded. That is, cyclodissipativeness with $Q < 0$ implies that the system has no purely imaginary poles. It may, however, have poles in the left or right half-planes. Even in this case, cyclodissipativeness does not imply stability.

For the circles in Theorem 28 to have any meaning, we must of course have $S^2 - QR > 0$. This is the assumption that we have imposed on the class of "interesting" (Q, S, R) triples. That assumption was not needed in the proof of Theorem 25, but it turns out be required anyway to get meaningful circle criteria. It may be verified that, in the absence of this assumption, the theorem ceases to say anything interesting about G. Depending on the sign of Q, either $M(j\omega) \geq 0$ can never be satisfied, or it is satisfied for every $G(j\omega)$. These remarks, however, apply only to the scalar case. When Q, S, and R are matrices, the situation becomes more complicated.

Let us now turn to dissipativeness. Our basic tool is Theorem 26.

THEOREM 29. *Let $S^2 - QR > 0$. Then the necessary and sufficient conditions for the scalar G to be (Q, S, R) dissipative are that $G(j\omega)$ satisfy the appropriate test of Theorem 28, together with the constraint*

(1) *If $Q > 0$: the number of counterclockwise encirclements of the critical disc by the graph of $G(j\omega)$, as ω varies from $-\infty$ to $+\infty$, is equal to the number of poles in Re $s \geq 0$ of $G(s)$;*
(2) *If $Q < 0$: $G(s)$ has no poles in Re $s \geq 0$;*
(3) *If $Q = 0$: $G(s)$ has no poles in Re $s > 0$, and at most simple poles on Re $s = 0$; and, if $j\omega_0$ is a pole of $G(s)$, then $S \lim_{s \to j\omega_0} (s - j\omega_0)G(s) > 0$.*

PROOF. The proofs follow from Theorem 26, by different choices of K in inequality 8.

(1) If $Q > 0$, choose $K = Q/S$, and note that this makes the number of encirclements of the critical disc precisely equal to the number of encirclements of the point $(-1/K, 0)$ in the Nyquist test.
(2) If $Q < 0$, choose $K = 0$.
(3) If $Q = 0$, choose $K = \varepsilon/S$ for small $\varepsilon > 0$, and note that Theorem 26 requires the closed loop poles to be in the open left half-plane for arbitrarily small ε. If there is an open-loop pole at $j\omega_0$, then standard root locus "angle of departure" arguments show that the pole must be simple and that $S \text{ Re} \lim_{s \to j\omega_0} (s - j\omega_0)G(s) > 0$. Finally, the "Re" qualifier can be dropped from this last condition, on the grounds that $M(j\omega) \geq 0$ for ω close to ω_0 implies that the residue is purely real. □

Notice that Theorem 27 was not used in deriving these results. An alternative approach would be to take Theorem 27 as a starting point, and then to prove Theorem 29 using the maximum modulus theorem of complex algebra. (For case 1, this involves arguments that almost identical with the standard proof of the Nyquist criterion.) The approach used here is somewhat simpler, since it takes the Nyquist theorem as a given result.

Case 2 of the theorem includes finite gain systems as a special case. If a system is to be $(-1, 0, -k^2)$ dissipative, then obviously the best choice for k is $\sup |G(j\omega)|$; accordingly, we can call this quantity the gain of G. Naturally, the gain is a meaningful quantity only when $G(s)$ has no poles in the closed right half-plane.

4. Graphical tests: the multivariable case

Let us turn now to the case where $G(s)$ is a matrix. In general, it is still possible to test for cyclodissipativeness via the condition $M(j\omega) \geq 0$; and one way to do this is to plot the eigenvalues of $M(j\omega)$ as a function of ω, and check that they remain nonnegative. This works if Q, S, and R are known. More often, the problem of interest is: given $G(s)$, find a suitable (Q, S, R) triple. This makes it more desirable to have tests that work directly from $G(j\omega)$.

Before proceeding, let us recall a simple fact from matrix algebra, a fact that is sometimes overlooked. The aforementioned test on the eigenvalues of $M(j\omega)$ can be justified only because $M(j\omega)$ is an Hermitian matrix; that is, the transpose of $M(j\omega)$ is equal to its complex conjugate. This implies, among other things, that the eigenvalues of $M(j\omega)$ are real. In the non-Hermitian case, matrix inequalities cannot be turned into eigenvalue inequalities; instead, one has to work with singular values. The singular values of a matrix A are the eigenvalues of $(A^*A)^{1/2}$. Notice that A^*A is always Hermitian, for any A.

Our first result is for the very special case in which $G(s)$ is a normal matrix. That is,

$$G(s) = V(s)\Lambda(s)V(s)^*$$

where

$$\Lambda(s) = \mathrm{diag}\{\lambda_1(s), \lambda_2(s), ..., \lambda_m(s)\}$$

and

$$V(s)^{-1} = V(s)^*$$

It is also assumed that $Q = qI$, $S = tI$, $R = rI$ for scalars q, t, and r. The lower case t is used to avoid confusion with the complex variable s.

THEOREM 30. *Let $G(s)$ be normal with eigenvalues $\lambda_i(s)$, and suppose that $t^2 > qr$. Then G is (qI, tI, rI) cyclodissipative or dissipative iff all $\lambda_i(s)$ satisfy the criteria of Theorem 28 or 29, respectively.*

PROOF. We have

$$\begin{aligned} M(s) &= G(s)^*QG(s) + G(s)^*S + S^TG(s) + R \\ &= V(s)[q\Lambda(s)^*\Lambda(s) + t\Lambda(s)^* + t\Lambda(s) + rI]V(s)^* \end{aligned}$$

and this is nonnegative definite iff

$$q\lambda_i(s)^*\lambda_i(s) + t\lambda_i(s)^* + t\lambda_i(s) + r \geq 0$$

for all i. □

The pole constraints of Theorem 29 apply, of course, to the poles of each $\lambda_i(s)$ separately. These conditions can if desired be expressed in terms of the poles of $G(s)$, but there is no computational advantage in doing so.

For the general case, we have to work in terms of singular values.

THEOREM 31. *Assuming $t^2 > qr$, the necessary and sufficient conditions for G to be (qI, tI, rI) cyclodissipative are:*

(1) If $q > 0$,
$$\sigma_{\min}(G(j\omega) + \frac{t}{q}I) \geq \frac{1}{q}\sqrt{t^2 - qr}$$
for all ω, where $\sigma_{\min}$ denotes the smallest singular value.
(2) If $q < 0$,
$$\sigma_{\max}(G(j\omega) + \frac{t}{q}I) \leq \frac{1}{|q|}\sqrt{t^2 - qr}$$
for all ω, where $\sigma_{\max}$ denotes the largest singular value.
(3) If $q = 0$,
$$t\lambda_{\min}(G(j\omega) + G(j\omega)^*) \geq -r$$
for all ω, where $\lambda_{\min}$ denotes the smallest eigenvalue.

Further, the system is (qI, tI, rI) dissipative iff in addition

(1) If $q > 0$, $qN(s) + tD(s)$ is nonsingular in Re $s > 0$, where $G(s) = D(s)^{-1}N(s)$;
(2) If $q < 0$, $G(s)$ has no poles in Re $s \geq 0$;
(3) If $q = 0$, $G(s)$ has no poles in Re $s > 0$ and at most simple poles on Re $s = 0$; and if $j\omega_0$ is a pole of $G(s)$, then the residue matrix $\lim_{s\to\omega_0}(s - j\omega_0)tG(s)$ is nonegative definite Hermitian.

The proof of this result is similar to that for the scalar case, and is therefore omitted.

Theorem 31 has two shortcomings. First, it assumes a restrictive form for the matrices Q, S, and R. Second, it requires one to work in terms of singular values and eigenvalues, giving limited insight into such practical questions as to how to tune the system in order to obtain the desired properties. Systems engineers are more likely to prefer tests based directly on the elements of $G(s)$; the remainder of this section is devoted to such tests. There is, however, a cost. Whereas Theorems

30 and 31 gave necessary and sufficient conditions for dissipativeness, the graphical tests to be given in the remainder of this section are sufficiency tests. In exchange for convenience, we lose some tightness on the bounds.

Most of the results, including Theorems 30 and 31, are for square $G(s)$. The non-square case — that is, the case of systems with unequal numbers of inputs and outputs — is to some extent still an open question. There is, however, one general result which can be helpful in the non-square case.

Theorem 32. *For any real matrix A of appropriate dimensions, the system with transfer function $AG(s)$ is (Q, S, R) (cyclo)dissipative iff the system with transfer function $G(s)$ is (A^TQA, A^TS, R) (cyclo)dissipative.*

The proof is obvious, from the definitions of dissipativeness and cyclodissipativeness.

While we are looking at transformations, it is also of interest to note the following result.

Theorem 33. *If system G is invertible, then it is (Q, S, R) (cyclo)dissipative iff G^{-1} is (R, S^T, Q) (cyclo)dissipative.*

Again, the proof is obvious, although the interpretation of this last theorem needs some care about the meaning of "invertible". If $G(s)$ is strictly proper, then its formal inverse $G(s)^{-1}$ is non-proper; that is, it has poles at infinity. A strict interpretation of the results of this chapter excludes such cases. In practice, a pole at infinity can be tolerated if the precaution used in applying the Nyquist criterion are applied: consider a D-shaped contour in the right half plane, and take the limit as this region expands to include the entire right half plane. For those of the preceding theorems that require checking a residue matrix in the case $q = 0$, the residue matrix for a pole at infinity is $\lim_{s\to\infty} \frac{1}{s} tG(s)$.

The utility of these last two "transformation" results lies in the fact that our later theorems give sufficient but not necessary conditions for dissipativeness. For example, some of the tests benefit from a diagonal dominance condition: the off-diagonal elements of $G(j\omega)$ should not be too large relative to the diagonal elements. It sometimes happens that $G(j\omega)^{-1}$ has this property even though $G(j\omega)$ does not, and in such cases the last theorem can be used.

Before proceeding to the main results, we need two preliminary lemmas.

Lemma 7. *Let Z be a complex square matrix with the properties*

$$\begin{aligned} \sum_j \frac{a_j}{a_i} |Z_{ij}| \le 1 \quad &\textit{for all } i \\ \sum_i \frac{a_i}{a_j} |Z_{ij}| \le 1 \quad &\textit{for all } j \end{aligned}$$

for some set of positive a_i. Then $I - Z^*Z \ge 0$.

PROOF. Let x be an arbitrary complex vector, and let $(Zx)_i$ denote the ith component of the vector Zx. Then

$$\begin{aligned}
|(Zx)_i|^2 &\leq \left(\sum_j |Z_{ij}|\,|x_j|\right)^2 \\
&= \left(\sum_j \left(\frac{a_j}{a_i}|Z_{ij}|\right)^{1/2}\left(\frac{a_i}{a_j}|Z_{ij}|\,|x_j|^2\right)^{1/2}\right)^2 \\
&\leq \left(\sum_j \frac{a_j}{a_i}|Z_{ij}|\right)\left(\sum_j \frac{a_i}{a_j}|Z_{ij}|\,|x_j|^2\right) \\
&\leq \sum_j \frac{a_i}{a_j}|Z_{ij}|\,|x_j|^2
\end{aligned}$$

From this it follows that

$$\sum_i |(Zx)_i|^2 \leq \sum_j |x_j|^2 \sum_i \frac{a_i}{aj}|Z_{ij}| \leq \sum_j |x_j|^2$$

This means that $x^*Z^*Zx \leq x^*x$, as desired. □

LEMMA 8. *Let Z be a complex square matrix whose diagonal entries are real, and suppose that there exist real $a_i > 0$ such that*

$$Z_{ii} - \frac{1}{2}\sum_{j\neq i}\frac{a_j}{a_i}\left|Z_{ij} + Z_{ji}^*\right| \geq 1$$

for all i. Then $Z + Z^* - 2I \geq 0$.

PROOF. Let $Y = \frac{1}{2}(Z + Z^*) - I$. The given constraints for Z map into the constraints

$$a_i Y_{ii} \geq \sum_{j\neq i} a_j\,|Y_{ij}|$$

Let x be an arbitrary complex vector. Then

$$x^*Yx = \sum_i\left(Y_{ii}\,|x_i|^2 + \sum_{j\neq i} Y_{ij}x_i^*x_j\right) \geq \sum_i\left(Y_{ii}\,|x_i|^2 - \sum_{j\neq i}|Y_{ij}|\,|x_i|\,|x_j|\right)$$

Now let $u_i = |x_i|/a_i$. This gives

$$x^*Yx \geq \sum_i\sum_{j\neq i} a_i a_j\,|Y_{ij}|\left(u_i^2 - u_i u_j\right)^2 = \sum_i\sum_{j>i} a_i a_j\,|Y_{ij}|\,(u_i - u_j)^2$$

where we have separated out the cases $j < i$ and $j > i$, and taken advantage of the fact that $|Y_{ji}| = |Y_{ji}|$. Clearly, then, $Y \geq 0$, which is the desired result. □

After these preliminaries, we can finally turn to graphical dissipativeness tests. It is convenient to adopt the following terminology. For a given transfer function matrix $G(s)$, the ith Nyquist band of radius $\delta_i(\omega)$ is the region swept out in the complex plane by a circle with centre $g_{ii}(j\omega)$ and radius $\delta_i(\omega)$, as ω varies from $-\infty$ to $+\infty$. In the special case where $G(s)$ is a diagonal matrix, these "Nyquist bands" will simply be Nyquist plots of the diagonal elements. The broadening of the plots caused by $\delta_i(\omega)$ is, in the general case, an allowance to take into account the off-diagonal elements of $G(j\omega)$.

THEOREM 34. *Let the ith Nyquist band of radius*

$$\max\left(\sum_{j\neq i}|g_{ij}(j\omega)|, \sum_{j\neq i}|g_{ji}(j\omega)|\right)$$

lie inside the circle with centre $(c_i,0)$ *and radius* ρ_i*, for each* i*. Define* $P = diag\{\rho_i\}$ *and* $C = diag\{c_i\}$*. Then* G *is* $(-P^{-1}, P^{-1}C, P - C^2P^{-1})$ *cyclodissipative, and dissipative if in addition* G *has no poles in the closed right half plane.*

PROOF. We have

$$|g_{ii}-c_i| + \sum_{j\neq i}|g_{ij}| \leq \rho_i$$

and

$$|g_{ii}-c_i| + \sum_{j\neq i}|g_{ji}| \leq \rho_i$$

for all i. Define $Z(j\omega) = P^{-1/2}(G(j\omega) - C)P^{-1/2}$; then it is easily shown that $Z(j\omega)$ satisfies the conditions of Lemma 7. The remaining details are obvious. □

THEOREM 35. *Let the ith Nyquist band of radius*

$$\frac{1}{2}\sum_{j\neq i}(|g_{ij}(j\omega)| + |g_{ji}(j\omega)|)$$

lie outside the circle with centre $(c_i,0)$ *and radius* ρ_i*, for each* i*. Define* $P = diag\{\rho_i\}$ *and* $C = diag\{c_i\}$*. Then* G *is* $(P^{-1}, -P^{-1}C, C^2P^{-1} - P)$ *cyclodissipative; and dissipative if in addition the sum of the net counterclockwise encirclements of the critical circles by the bands is equal to the number of poles of* $G(s)$ *in the closed right half plane.*

PROOF. Let

$$V = \text{diag}\left\{\frac{g_{ii}-c_i}{|g_{ii}-c_i|}\right\}$$

and

$$Z = V^*P^{-1/2}(G-C)P^{-1/2}$$

Then it may be shown from Lemma 8 that $Z + Z^* - 2I \geq 0$, and then that

$$Z^*Z - I = (Z-I)^*(Z-I) + Z + Z^* - 2I \geq 0$$

The remaining details are obvious. □

Notice that Theorem 34 covers the case where all the Nyquist bands lie inside critical circles, and that Theorem 35 covers the case where they lie outside critical circles. It would be useful to have a result for the more general case, where the ith Nyquist band lies outside a circle for some i, and inside a circle for the remaining i, leading to (Q,S,R) dissipativeness where Q is sign indefinite. This is still an open problem.

The next result, for $(0,S,R)$ dissipativeness, is in the same family of tests, but it is interesting to see that the bands used can be a little narrower.

THEOREM 36. *Let* $S = diag\{\sigma_i\}$*, where each* σ_i *can be +1 or -1. Suppose that the ith Nyquist band of radius* $\frac{1}{2}\sum_{j\neq i}\frac{a_j}{a_i}\left|\sigma_i g_{ij} + \sigma_j g_{ij}^*\right|$*, for arbitrary real constants* $a_i > 0$*, lies to the left (if* $\sigma_i < 0$*) or to the right (if* $\sigma_i > 0$*) of the line Re* $s = b_i$*, and let* $B = diag\{b_i\}$*. Then* G *is* $(0, S, -2BS)$ *cyclodissipative; and dissipative if in addition* $G(s)$ *has no poles in the closed right half plane.*

Proof. The bounds are of the form

$$\sigma_i \mathrm{Re}\ g_{ii} \geq \sigma_i b_i + \frac{1}{2}\sum_{j\neq i}\frac{a_j}{a_i}\left|\sigma_i g_{ij} + \sigma_j g_{ji}^*\right|$$

which reduces to

$$a_i Y_{ii} \geq \sum_{j\neq i} a_j \left|Y_{ij}\right|$$

where $Y = SG + G^*S - 2BS$. One can then argue, as in Lemma 8, that $Y \geq 0$, which gives the desired result. □

To apply this result in practice, the easiest approach is first to choose $a_i = 1$ for all i, and use the conservative estimate $\frac{1}{2}\sum_{j\neq i}(|g_{ij}| + |g_{ji}|)$ for the radius of the ith band. This allows a decision to be made on the sign of each σ_i. Narrower bands can be drawn when the σ_i are all known, to allow the b_i to be measured. If desired, one can then experiment with different a_i to get even tighter estimates.

The last three theorems may be interpreted as saying that G inherits the dissipativeness properties of its diagonal entries, with a correction to take into account the influence of the off-diagonal terms. Clearly, these tests will give good results when the off-diagonal terms are small with respect to the diagonal terms, but could be overly conservative otherwise. Let us now consider an alternative approach, which in some cases is less restrictive in the constraints it places on the $g_{ij}(j\omega)$.

In each of the tests that follow, it is assumed that constant matrices B and C have been selected such that

$$|g_{ij}(j\omega) - c_{ij}| \leq b_{ij}$$

for all ω, and all i and j with $i \neq j$. This is itself a circle test, and it should be clear how the "best" b_{ij} and c_{ij} can be selected graphically. The diagonal entries b_{ii} and c_{ii} are unspecified so far, but will be given in the theorem statements that follow.

The results rely on the properties of M-matrices. Briefly, a real square matrix is an M-matrix if its off-diagonal elements are all nonpositive and its leading principal minors are positive. More details are given in the Appendix. For our present purposes, we need the following results from the Appendix.

(1) Let B be a real square matrix all of whose entries are nonnegative, and let K be a diagonal matrix whose diagonal entries are sufficiently large to make $K - B$ an M-matrix. Then there exists a positive definite diagonal matrix D such that $B^T DB \leq DK^2$.
(2) Let $(-B)$ be an M-matrix. Then, for any positive definite diagonal K such that $(-B - K)$ is also an M-matrix, there exists a positive definite diagonal matrix D such that $B^T DB \geq DK^2$. Further, for the same D and K we have
$$\tilde{B}^T D\tilde{B} \geq DK^2$$
for any $\tilde{B} = [\tilde{b}_{ij}]$ that satisfies
$$\begin{aligned} \tilde{b}_{ii} \leq b_{ii} < 0 &\qquad \text{for all } i \\ 0 < \tilde{b}_{ij} \leq b_{ij} &\qquad \text{for all } j \neq i \end{aligned}$$
(3) If M is an M-matrix, then there exists a positive definite diagonal D such that $DM + M^T D > 0$.

A further relevant point is that the matrices D and K in these assertions are easy to compute. For further details, see the Appendix.

Notice, by the way, that if we did not require D to be diagonal than the M-matrix conditions could be replaced by conditions on the eigenvalues of the matrices in question. It turns out, however, that the proofs of the following three theorems rely critically on the fact that D is diagonal.

THEOREM 37. *Suppose that the Nyquist plot of $g_{ij}(j\omega)$ lies inside the circle with centre $(c_{ij}, 0)$ and radius b_{ij}, for all i and j. Then G is $(-D, DC, DK^2 - C^T DC)$ cyclodissipative, where K and D are diagonal positive definite matrices, chosen such that $K - B$ is an M-matrix and D is a solution of $B^T DB \leq DK^2$. It is dissipative if in addition $G(s)$ has no poles in Re $s \geq 0$.*

PROOF. The circle conditions give

$$\|y_i - (Cu)_i\|_T \leq \sum_j b_{ij} \|u_j\|_T$$

for all i, and therefore

$$\sum_i d_i \|y_i - (Cu)_i\|_T^2 \leq \sum_i d_i \left(\sum_j b_{ij} \|u_j\|_T \right)^2$$

for any set of $d_i > 0$. The left side of this inequality can easily be written as an inner product, but the right side can not. To get around this problem, let $v_i = \|u_i\|_T$. Then

$$\begin{aligned} \langle y - Cu, D(y - Cu) \rangle_T &\leq \sum_i d_i (Bv)_i^2 \\ &= v^T B^T DBv \\ &\leq v^T DK^2 v \\ &= \sum_i d_i k_i^2 \|u_i\|_T^2 \\ &= \left\langle u, DK^2 u \right\rangle_T \end{aligned}$$

This gives the desired result. □

THEOREM 38. *Suppose that the Nyquist plot of $g_{ii}(j\omega)$ lies outside the circle with centre $(c_{ii}, 0)$ and radius b_{ii}, for all i, and for all $j \neq i$ the Nyquist plot of $g_{ij}(j\omega)$ lies inside the circle with centre $(c_{ij}, 0)$ and radius b_{ij}. Then G is $(D, -DC, C^T DC - DK^2)$ cyclodissipative, where K is a diagonal positive definite matrix chosen such that $(-B-K)$ is an M-matrix, and D is a diagonal positive definite matrix chosen such that D is a solution of $B^T DB \leq DK^2$. It is dissipative if in addition $g_{ij}(s)$ has no poles in Re $s \geq 0$ for $j \neq i$, and the number of anti-clockwise encirclements of the critical disk by $g_{ii}(s)$ is equal to the number of right half plane poles of $g_{ii}(s)$.*

PROOF. The circle conditions now give

$$\begin{aligned} \|y_i - (Cu)_i\|_T &= \left\| (g_{ii} - c_{ii})u_i + \sum_{j \neq i} (g_{ij} - c_{ij})u_j \right\|_T \\ &\geq \|(g_{ii} - c_{ii})u_i\|_T - \sum_{j \neq i} \|(g_{ij} - c_{ij})u_j\|_T \\ &\geq -b_{ii} \|u_i\|_T - \sum_{j \neq i} b_{ij} \|u_j\|_T \end{aligned}$$

We cannot square this inequality, as in the previous theorem, because the right side could be positive or negative, depending on $\|u\|_T$. (Notice that $b_{ii} < 0$, and $b_{ij} \geq 0$

for $j \neq i$.) To get around this problem, we can replace the inequality by

$$\sum_i \|y_i - (Cu)_i\|_T^2 \geq -\sum_j \tilde{b}_{ij} \|u_j\|_T \geq 0$$

where $\tilde{b}_{ij} = b_{ij}$ whenever the original right side is positive; and otherwise $\tilde{b}_{ij}$ is a modification of b_{ij} just sufficient to bring the right side up to zero. A little thought will show that this can always be done consistently with the earlier stated conditions that ensure $\tilde{B}^T D\tilde{B} \geq DK^2$. The remainder of the proof is almost identical with the proof of Theorem 37. □

THEOREM 39. *Suppose g_{ii} is $(0, t_i, r_i)$ (cyclo)dissipative, and the off-diagonal elements satisfy $|g_{ij}(j\omega) - c_{ij}| \leq b_{ij}$ for $j \neq i$. Let $c_{ii} = 0$ for all i. Choose $K = diag\{k_i\}$ such that the matrix M with elements*

$$\begin{aligned} m_{ii} &= k_i - r_i \\ m_{ij} &= -|t_i|\, b_{ij} \text{ for } j \neq i \end{aligned}$$

is an M-matrix, and let D be a diagonal positive definite solution of $DM + M^T D > 0$. Then G is $(0, DT, DK - TDC - C^T DT)$ (cyclo)dissipative, where $T = diag\{t_i\}$.

PROOF. The dissipativeness conditions imply

$$\begin{aligned} 2t_i \langle y_i - (Cu)_i, u_i\rangle_T &\geq -r_i \|u_i\|_T^2 + \sum_{j\neq i} 2t_i \langle (g_{ij} - c_{ij})u_j, u_i\rangle_T \\ &\geq -r_i \|u_i\|_T^2 - \sum_{j\neq i} 2\,|t_i|\,|b_{ij}|\,\|u_i\|_T\,\|u_j\|_T \end{aligned}$$

Let $v_i = \|u_i\|_T$. Then

$$2t_i \langle y_i - (Cu)_i, u_i\rangle_T + k_i \|u_i\|_T^2 \geq \sum_j m_{ij} v_i v_j$$

and therefore

$$\sum_j 2d_i t_i \langle y_i - (Cu)_i, u_i\rangle_T + \sum_j d_i k_i \langle u_i, u_i\rangle_T \geq v^T (DM + M^T D)v \geq 0$$

The remaining details are obvious. □

In effect, Theorems 37 to 39 view the system as an interconnection of scalar subsystems, and show that, under certain conditions, dissipativeness of the subsystems implies dissipativeness of the overall system. In principle, it should be possible to mix the conditions of these theorems, for example by allowing some of the g_{ii} to lie outside circles and others to lie inside circles. The derivation of dissipativeness tests of that type is still an open problem.

A natural question is whether Theorems 37 to 39 are more or less conservative than Theorems 34 to 36. There seems to be no clear-cut answer. A shortcoming of the M-matrix tests is that they require the condition $|g_{ij}(j\omega) - c_{ij}| \leq b_{ij}$ to hold independently for each $j \neq i$. The constants b_{ij} and c_{ij} cannot depend on ω, and there is no trade-off possible if, for example, the g_{ij} are large in different frequency ranges for different j. In contrast Theorem 36, in particular, allows frequency-dependent cancellations between g_{ij} and g_{ji}^*. On the other hand, the M-matrix conditions themselves are generally less restrictive than the conditions of Theorems 34 to 36. What this means in practice is that, for any given example, it is worth trying both types of test to see which one gives the best results.

To close this section, let us consider a result which is closely related to Theorem 35. The proof is not easy, but the result is of considerable historical importance; it is, in effect, a mapping into our terminology of the first known multivariable circle criterion.

THEOREM 40. *Let* $\Theta = diag\{\theta_i\}$ *satisfy* $\sum_i \theta_i^{-2} \leq 1$, *and suppose there exist* $Z = diag\{\zeta_i\}$ *and* $H = diag\{\eta_i\}$ *such that the Nyquist plots satisfy either*

$$|\zeta_i + g_{ii}| - \sum_{j \neq i} |g_{ij}| > \theta_i \eta_i \text{ for all } i$$

or

$$|\zeta_i + g_{ii}| - \sum_{j \neq i} |g_{ji}| > \theta_i \eta_i \text{ for all } i$$

Then G *is* $(D^2, D^2 Z, D^2(Z^2 - H^2))$ *cyclodissipative, and dissipative with the addition of the usual encirclement conditions, where* D *is* $(\Theta H)^{-1}$ *in the case of the first inequality being satisfied, or* Θ *in the second case.*

PROOF. Let $A^{-1} = Z - H$ and $B^{-1} = Z + H$. Rosenbrock [**Ros70**] showed that the given conditions imply that Y is positive real, where the matrix Y is defined as $Y = (B^{-1} + \Theta^{-1} G \Theta)^{-1}(A^{-1} + \Theta^{-1} G \Theta)$ in the first case, and $Y = (A^{-1} + \Theta G \Theta^{-1})(B^{-1} + \Theta G \Theta^{-1})^{-1}$ in the second case. In that second case, the positive real condition $Y + Y^* \geq 0$ reduces immediately to the desired dissipativeness condition. The first case is more complicated, but may be handled by a technique due to Araki [**Ara75**]. Let $X = (Y + I)^{-1}(Y - I)$; then the condition $Y + Y^* \geq 0$ implies the bounded real condition $I - X^* X \geq 0$, which in turn reduces to the given dissipativeness condition. □

By analogy with the earlier results in this section, one might expect two more results along these lines. That is, similar criteria for (Q, S, R) dissipativeness where $Q < 0$ and $Q = 0$. It is easy to conjecture the probable forms of such conditions, but so far no proofs are known. Another open problem is to find a more direct proof of Theorem 40.

5. Results using multipliers

A standard trick in stability testing is to introduce extraneous dynamics into the system, in the hope of adding extra flexibility to the stability criteria. A typical situation is shown in figures 1 and 2. Figure 1 is a feedback system whose stability we wish to investigate. For concreteness, it can be supposed that G is a linear system characterised by a transfer function $G(s)$, and that N is a memoryless nonlinearity. Figure 2 shows an augmented system, in which a linear "multiplier" Z, with transfer function $Z(s)$, is inserted in the loop, with Z^{-1} also inserted to keep the loop gain unchanged. Clearly, it is not hard to establish conditions under which stability of the augmented system implies stability of the original system.

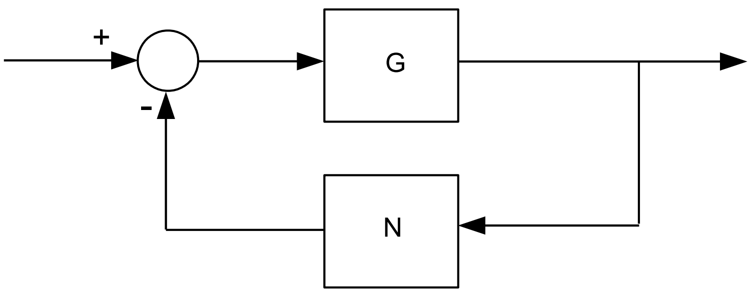

FIGURE 1. A feedback system

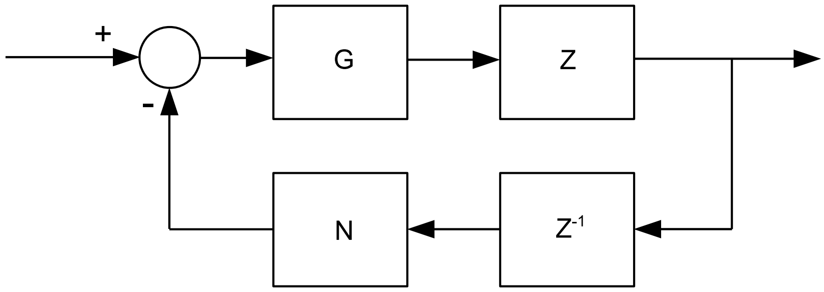

FIGURE 2. The augmented system with multiplier **Z**

A similar technique may be used for more general multi-loop systems. The main requirement is that, whenever a multiplier is inserted into the block diagram, its inverse must also be inserted at the appropriate points.

As we have seen in earlier chapters, a stability analysis can be carried out by finding the dissipativeness parameters of subsystems. That means that we want to find the dissipativeness parameters of ZG and NZ^{-1}. This will, one hopes, lead to less restrictive conditions than for the case $Z = I$.

The analysis of the nonlinear system NZ^{-1} will be left to the next chapter. For now, we reassure the reader that there are indeed interesting classes of N and Z for which useful results can be obtained. The topic of the present section is to obtain dissipativeness conditions for the linear system ZG. In connection with this problem, one can make several preliminary observations:

(1) If $Z(s)$ is known, then of course one can simply compute the transfer function of $F(s) = Z(s)G(s)$, and then use the methods given earlier. Normally, however, $Z(s)$ is not known, and part of the problem is to find a suitable $Z(s)$. This means that we need tests that are based on $G(s)$ alone.

(2) Because we are looking for *graphical* tests, a standing assumption throughout this section will be that $G(s)$ is a scalar transfer function. Although some results are known for the multivariable case, they are not easy to express in graphical form.

(3) It is sometimes necessary to assume that there are no pole-zero cancellations in forming the product $Z(s)G(s)$. Actually, no such assumption is needed for the dissipativeness analysis itself, but it might be needed in concluding that dissipativeness implies stability. The rule is not invariable — a pole-zero cancellation in the left half plane can often be tolerated — and we recommend that each case be treated separately on its merits.

(4) For the most part, only very simple forms of $Z(s)$ are worth considering. For a complicated $Z(s)$, it becomes too difficult to deduce the properties of $Z(s)G(s)$ from those of $G(s)$.

(5) The circle criteria of earlier sections were based on the Nyquist plot of $G(j\omega)$, a graph in which the horizontal axis is Re $G(j\omega)$ and the vertical axis is Im $G(j\omega)$. Most of the tests in this section use a *Popov plot*, where the horizontal axis is Re $G(j\omega)$ as before, but the vertical axis is ωIm $G(j\omega)$.

We shall adopt the following notation. With $F(s) = Z(s)G(s)$, let $F_r =$ Re $F(j\omega)$ and $F_i =$ Im $F(j\omega)$. Likewise, let $G_r =$ Re $G(j\omega)$ and $G_i =$ Im $G(j\omega)$.

The quantity we want to test is

$$M(j\omega) = R + S^T F(j\omega) + F(j\omega)^* S + F(j\omega)^* Q F(j\omega)$$

which simplifies to

$$M(j\omega) = R + 2SF_r + Q(F_r^2 + F_i^2)$$

because all quantities are scalars. For useful graphical criteria, the condition $M(j\omega) \geq 0$ must be turned into an inequality involving G_r and G_i.

Of course, dissipativeness also imposes conditions on the poles of $F(s)$ in Re $s \geq 0$, and in some cases residue conditions where there are imaginary poles. To keep the discussion simple, it will be assumed throughout this section that $F(s)$ has no poles in Re $s \geq 0$, and no poles at infinity. The details of the more general case can easily be filled in by the reader.

Let us begin with the very simplest case: $Z(s) = s$. This gives $F_r = -\omega G_i$, $F_i = \omega G_r$, and therefore

$$M(j\omega) = R - 2S\omega G_i + \omega^2 Q(G_r^2 + G_i^2)$$

If $Q = 0$, then $M(j\omega) \geq 0$ is equivalent to $\omega G_i \geq \frac{R}{2S}$ or $\omega G_i \leq \frac{R}{2S}$, depending on the sign of S. Thus, the required condition is that the Popov plot lie above or below (depending on the sign of S) a horizontal straight line. This is very different from the familiar passivity-like criteria, where the boundary is a vertical straight line.

When $Q \neq 0$, the appropriate graph is a graph of ωG_i against ωG_r, and the condition is that this graph lie inside or outside a circle, depending on the sign of Q.

The foregoing results are fairly elementary, but they do illustrate the basics of the multiplier method. Let us now turn to a more interesting case: a multiplier of the form $Z(s) = 1 + \alpha s$, where α is a (so far unspecified) constant. Here we have $F_r = G_r - \alpha\omega G_i$, $F_i = \alpha\omega G_r + G_i$, so that the quantity of interest is

$$M(j\omega) = R + 2S(G_r - \alpha\omega G_i) + Q(G_r^2 + G_i^2 + \alpha^2\omega^2 G_r^2 + \alpha^2\omega^2 G_i^2)$$

The case $Q = 0$, which corresponds to the standard Popov criterion, is particularly important.

THEOREM 41. *For the multiplier $Z(s) = 1 + \alpha s$, the system ZG is $(0, S, R)$ cyclodissipative if the Popov plot of $G(j\omega)$ lies to the right, if $S > 0$, or to the left, if $S < 0$, of the straight line of slope $1/\alpha$ which passes through the point $(-\frac{R}{2S}, 0)$.*

This is illustrated in Figure 3, for the case where S, R, and α are all positive. This corresponds to the standard form of the Popov criterion. Notice that Theorem 41, unlike the standard Popov criterion, puts no sign constraints on these constants. (We may, however, discover that α has to be sign-constrained once we start investigating the nonlinear part of the system.) It can be seen that the line of Theorem 41 crosses the real axis at precisely the same point as the line in the third part of Theorem 28.

When $Q \neq 0$, $M(j\omega)$ depends on the four variables G_r, G_i, ωG_r, and ωG_i. To obtain graphical criteria on a two-dimensional plot, it is necessary to eliminate two of these variables. For the case $Q > 0$, there are two approaches that give useful results.

THEOREM 42. *With $Q > 0$, let p_1 and p_2 denote the two points on the real axis where*

$$x = -\frac{S}{Q} \pm \frac{\sqrt{S^2 - QR}}{Q}$$

Then there exists a multiplier $Z(s) = 1 + \alpha s$ such that ZG is (Q, S, R) cyclodissipative if the Popov plot of $G(j\omega)$ lies outside either of the two regions:

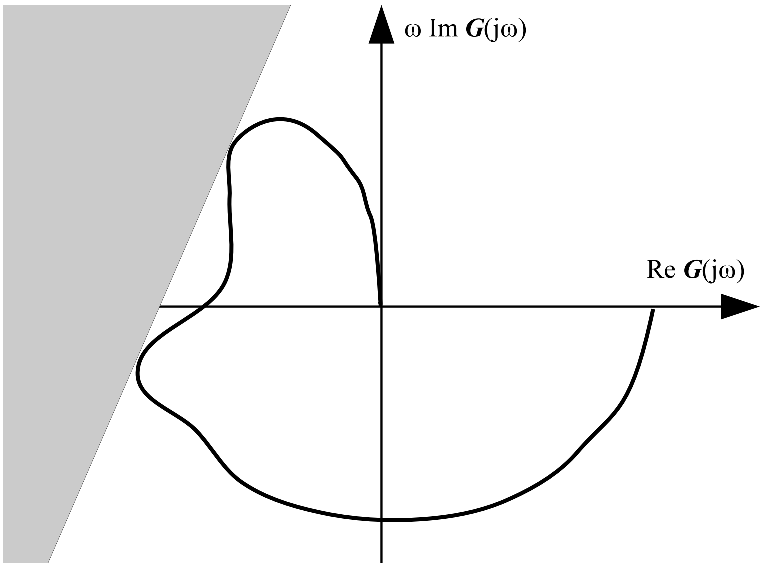

FIGURE 3. The Popov criterion

(1) *the ellipse with boundary*

$$\left(x+\frac{S}{Q}\right)^2+\alpha^2\left(y-\frac{S}{\alpha Q}\right)^2=\frac{2S^2-RQ}{Q^2}$$

which passes through the points p_1 *and* p_2*; or*

(2) *the region bounded by the parallel straight lines, of slope* $1/\alpha$*, passing through the points* p_1 *and* p_2*.*

PROOF. For the first condition, notice that

$$M(j\omega)\geq R+2S(x-\alpha y)+Q(x^2+\alpha^2y^2)$$

where $x = \text{Re } G(j\omega)$ and $y = \omega\text{Im } G(j\omega)$. The right side of this inequality is zero on an ellipse, with centre $\left(-\frac{S}{Q}, \frac{S}{\alpha Q}\right)$ which passes through the points p_1 and p_2, and positive outside the ellipse. For the second condition, we may use the alternative inequality

$$M(j\omega)\geq R+2S(x-\alpha y)+Q(x-\alpha y)^2$$

which shows that $M(j\omega)$ is nonnegative whenever $(x-\alpha y)$ lies outside the bounds described by two parallel lines passing through p_1 and p_2. □

The two cases are illustrated in Figure 4, for the case where both p_1 and p_2 lie to the left of the origin. It is also possible that one of both of them lies to the right of the origin. The ellipse bound is usually less restrictive than the straight line bounds, but there could be some transfer functions where the opposite is true. The most interesting feature of Theorem 42 is that the points p_1 and p_2 are precisely the points cut by the circle in the standard circle criterion. Theorem 42 is in a sense less restrictive than the usual circle criterion, because it allows the centre of the ellipse to be displaced from the real axis. (This does, however, increase the area of the ellipse.) Unfortunately a direct comparison with the circle criterion would be unrealistic, primarily because the dissipativeness conditions on the nonlinear block NZ^{-1}, to be derived in the next chapter, are typically more stringent than those on N alone.

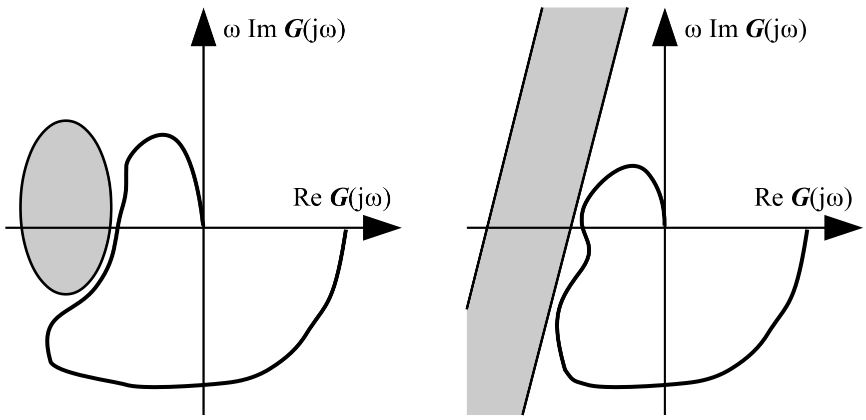

FIGURE 4. Bounds on the Popov plot for Theorem 42

The results for $Q < 0$ are less satisfactory, and not worth presenting here. Basically, the conditions are that the Popov plot lie inside a parabola, that $G(s)$ have no poles in Re $s \geq 0$, and that $|G(j\omega)| < K/\omega$ for some constant K. This last condition simply says that $G(s)$ must have more poles than zeros. Unfortunately the location of the parabola depends on K, in such a way that the conditions are unreasonably restrictive unless K is small.

An alternative way to introduce a multiplier is shown in Figure 5, where h is a real constant gain. It is easy to show that the re-drawn block diagram is equivalent, given suitable initial conditions for the Z and Z^{-1} subsystems, to the original feedback loop. Thus, it is of interest to derive dissipativeness conditions for the linear system with transfer function $ZG/(1 + hG)$.

It is, incidentally, very easy to show that $G/(1+hG)$ is (Q, S, R) dissipative iff G is $(Q + 2hS + h^2R, S + hR, R)$ dissipative. This means that if Z were inside the inner loop — that is, if the subsystem of interest were $ZG/(1+hZG)$ — we would

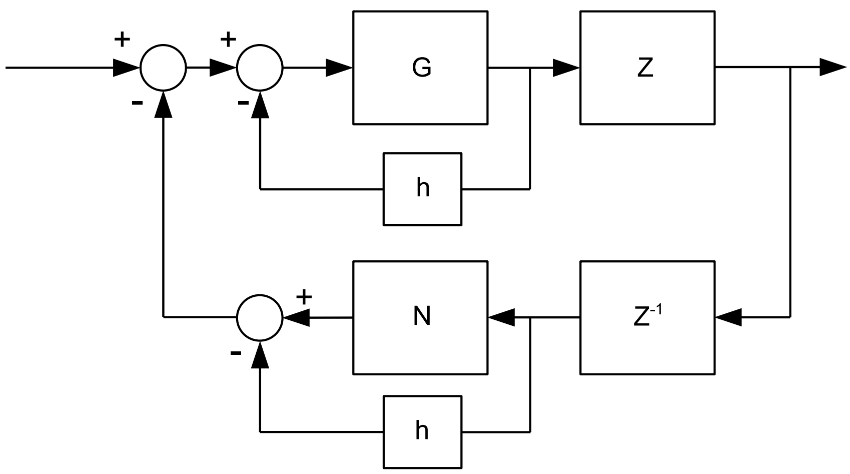

FIGURE 5. An alternative augmentation of the system

not have a new situation. The present reformulation is interesting only because Z does not occur in the denominator of $ZG/(1 + hG)$.

THEOREM 43. *If $h(2S + hR) \geq 0$, the linear system with transfer function $(1+\alpha s)G(s)/(1+hG(s))$ is $(0, S, R)$ cyclodissipative if the Popov plot of $G(j\omega)$ lies in the region defined by*

$$2\alpha Sy \leq (1 + hx)(R + (2S + hR)x)$$

PROOF. Letting $x = G_r = \text{Re } G(j\omega)$, $G_i = \text{Im } G(j\omega)$, and $y = \omega G_i$, a simple calculation shows that

$$|1 + hG(j\omega)|^2 M(j\omega) = h(2S + hR)G_i^2 + (1 + hx)(R + (2S + hR)x) - 2\alpha Sy$$

which is nonnegative under the given conditions. □

The boundary of the inequality in the theorem is a downward-pointing parabola. Depending on the sign of αS, the Popov plot must lie above or below this parabola. If no multiplier had been used — and if h had been known in advance — then G could have been tested for $(2hS + h^2R, S + hR, R)$ cyclodissipativeness, using the circle criterion of Theorem 28. It is interesting to note that the circle and the parabola cross the x axis at precisely the same two points. Those same crossing points also occur in the ellipse test of Theorem 42; and in yet another ellipse test, which follows.

THEOREM 44. *Let $Q_1 = Q + 2hS + h^2R$ and $S_1 = S + hR$. If $Q \geq 0$ and $Q_1 \geq 0$, then the linear system with transfer function $(1 + \alpha s)G(s)/(1 + hG(s))$ is (Q, S, R) cyclodissipative if the Popov plot of $G(j\omega)$ lies in the region defined by*

$$Q_1x^2 + 2S_1x + \alpha^2Qy^2 - 2\alpha Sy + R \geq 0$$

PROOF. Almost identical to the proof of Theorem 43. □

When Q and Q_1 are both positive, this inequality describes the exterior of an ellipse, and the conclusion of Theorem 44 is very similar to that of Theorem 42. There is a difference, however, in the relationship between the centre of the ellipse and its eccentricity.

When $Q = 0$, the result collapses back to the result of Theorem 43. When Q is positive but $Q_1 = 0$, the ellipse turns into a sideways-pointing parabola. When Q and Q_1 are both zero, the boundary is a straight line.

As a further illustration of what is possible, let us consider a more complicated multiplier.

THEOREM 45. *Let $Q_1 = Q + 2hS + h^2R$, $S_1 = S + hR$, and $Q_2 = \alpha^2Q + 2\alpha\beta hS + \beta^2h^2R$. Then the linear system with transfer function*

$$G_1(s) = \frac{1 + \alpha s}{1 + \beta s}\frac{G(s)}{1 + hG(s)}$$

is (Q, S, R) cyclodissipative if $Q_1 \geq 0$ and the Popov plot of $G(j\omega)$ lies in the region defined by the two bounds

$$Q_2x^2 + 2\beta(\alpha S + \beta hR)x + \beta^2R \geq 0$$
$$Q_1x^2 + Q_2y^2 + 2S_1x - 2S(\alpha - \beta)y + R \geq 0$$

PROOF. Let $M(j\omega) = R + 2S\text{Re } G_1(j\omega) + Q\,|G_1(j\omega)|^2$. Expanding out this expression, the result is

$$\begin{aligned}
&(1 + \beta^2\omega^2)\,|1 + hG(j\omega)|^2\, M(j\omega) \\
&\quad= \quad Q_1G_r^2 + Q_2(\omega G_i)^2 + 2S_1G_r - 2S(\alpha - \beta)\omega G_i + R \\
&\qquad\quad +Q_1G_i^2 + \omega^2(Q_2G_r^2 + 2\beta(\alpha S + \beta hR)(G_r + \beta^2R)
\end{aligned}$$

which is nonnegative if the conditions of the theorem are satisfied. □

The first inequality here describes a region bounded by a pair of vertical lines. The plot must lie between the lines if $Q_2 < 0$, outside them if $Q_2 > 0$. (If $Q_2 = 0$, we have only a single line.) The second inequality describes a region whose boundary is a conic section: an ellipse, a parabola, a pair of hyperbolae, or one or two straight lines, depending on the signs of the coefficients. In most cases the two regions overlap to some extent. The overall frequency domain condition is sometimes quite restrictive, and sometimes not, depending on the combination of parameters α, β, and so on.

By a regrouping of the terms in the proof of this theorem, a variant of Theorem 45 can be obtained in which the inequalities are conditions on the Nyquist plot rather than the Popov plot. The boundaries in the Nyquist plane are a pair of conic sections; the details are left to the reader.

Evidently, more complicated multipliers are going to lead to more complicated restrictions in the frequency domain, and it is doubtful whether it is worth looking at more general cases. An exception is the class of RL and RC multipliers, defined as follows.

Definition 19. *$Z(s)$ is in class RL if it has the form*

$$Z(s) = \prod_{n=0}^{N} \frac{s - \alpha_n}{s - \beta_n}$$

for some N, where the constants α_n and β_n are real and satisfy $0 < \alpha_0 < \beta_0 < \alpha_1 < \beta_1 \ldots$.

Definition 20. *$Z(s)$ is in class RC if $Z(s)^{-1}$ is in class RL.*

That is, the poles and zeros of an RL or RC function are real, and alternate along the negative real axis. The relationship between these definitions and the properties of RL and RC circuits should be obvious. To obtain useful results, we deliberate exclude the possibilities of poles or zeros at the origin or at infinity.

Lemma 9. *For any given $0 < a < b < \infty$, any $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, and any $\varepsilon > 0$, there exists an RL or RC function such that*

$$|phase\ Z(j\omega) - \theta| < \varepsilon$$

for all ω in $[a, b]$.

The proof of this lemma may be found in Cho and Narendra [**CN68**]. Briefly, an RL function will work if $\theta \geq 0$, and an RC function is appropriate if $\theta < 0$. Notice that we say nothing about how to construct $Z(s)$, which will be of high order if the bounds are tight. Lemma 9 will be applied in such a way that the actual $Z(s)$ is of no importance; it is sufficient to know that a suitable multiplier exists.

Lemma 10. *Suppose that the Nyquist plot of $G(j\omega)$ lies to the right, if $S > 0$, or to the left, if $S < 0$, of a straight line passing through the origin, and that it does not touch this line for any finite ω. Then there exists an RL or RC multiplier Z such that ZG is $(0, S, 0)$ cyclodissipative.*

Outline proof. The graphical condition implies that the phase of $SG(j\omega)$ lies in the range $(-\pi/2 - \theta, \pi/2 - \theta)$, where θ is the angle that the straight line makes with the imaginary axis. From Lemma 9, there exists a suitable $Z(s)$ such that the phase of $SZ(j\omega)G(j\omega)$ lies in the range $(-\pi/2, \pi/2)$ over a finite frequency range. The fact that $G(s)$ is real rational places restrictions on the phase of $G(j\omega)$ as $\omega \to 0$ or $\omega \to \infty$, so with care this argument can be extended to all ω.

Lemma 10 can be used to derive a family of stability criteria known as the *off-axis circle criteria*. Suppose that the Nyquist plot of $G(j\omega)$ lies inside a circle which intersects the real axis at the points $(-1/K_2, 0)$ and $(-1/K_1, 0)$. Unlike the standard circle criteria, the centre of the circle can be anywhere in the complex plane. Using the fact that bilinear transformations map circles to circles, with straight lines as the degenerate case, it can be shown that

$$\frac{1 + K_2 G(j\omega)}{1 + K_1 G(j\omega)}$$

satisfies the conditions of Lemma 10.

6. Discrete-time systems

Many of the results of this chapter apply virtually without change to a system described by difference equations

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned}$$

where now we use ℓ_2 rather than L_2 signal spaces. The usual transfer function description of this system is via the z transform

$$G(z) = D + C(zI - A)^{-1}B$$

The key change is to replace the half-plane Re $s \geq 0$ by the region $|z| \geq 1$. Correspondingly, plots of $G(j\omega)$ for real ω must be replaced by plots of $G(e^{j\theta})$, as θ varies from 0 to 2π; and statements about poles in the right half-plane must be replaced by statements about poles outside or on the unit circle.

With these changes, all results of the first half of this chapter — including the multivariable tests — are applicable to the discrete-time case virtually without modification. Theorem 29 needs a new proof, which the reader can easily supply. For the remaining theorems and lemmas, even the proofs carry over with negligible change.

It is only when multipliers are introduced that the discrete-time results differ significantly from their continuous-time counterparts. The following simple result illustrates the way in which the differences appear.

THEOREM 46. *The linear discrete-time system with transfer function $G_1(z) = (1 + \alpha z)G(z)$ is $(0.S, R)$ dissipative if the graph of $y = Re\ e^{j\theta}G(e^{j\theta})$ against $x = Re\ G(e^{j\theta})$, as θ varies from 0 to 2π, lies in the region*

$$R + 2Sx + 2\alpha Sy \geq 0$$

PROOF. We must evaluate

$$M(z) = R + 2S\mathrm{Re}\ G_1(z) + Q\,|G_1(z)|^2$$

as z moves around the circle $z = e^{j\theta}$. Letting $G(e^{j\theta}) = G_r + jG_i$, the result is

$$\begin{aligned} M(e^{j\theta}) &= R + 2SG_r + 2\alpha S(G_r \cos\theta - G_i \sin\theta) \\ &= R + 2Sx + 2\alpha Sy \end{aligned}$$

with the appropriate definition of x and y. □

The inequality in this condition is formally the same as the inequality required in the continuous-time Popov criterion. However the question of turning this into a stability criterion, and indeed the whole issue of multipliers for discrete-time systems, remains very much an open question.

7. Notes and references

Many of the results of this chapter are previously unpublished. Those which are not new have been presented in a very different order from their historical development; we have taken advantage of a great deal of hindsight to unify the results and simplify the proofs.

The first general graphical test was that of Popov ([**Pop61**], but more readily available in [**Pop73**]); it corresponds to Theorem 41. The next was the scalar circle criterion, Theorem 29, due to Sandberg [**San64**] and Zames [**Zam66**]. (Both Sandberg and Zames used a more general setting that included multivariable systems; but only in the scalar case were their results explicitly graphical.) Numerous variants of these criteria subsequently appeared. Our Theorems 42 to 45 were inspired by the parabola criterion of Bergen and Shapiro [**BS67**]. The off-axis circle criterion was due to Cho and Narendra [**CN68**].

Rosenbrock [**Ros70**] produced the first multivariable circle criterion, Theorem 40. Theorem 35 is essentially equivalent to a result by Cook [**Coo74**]. Theorems 37 to 39 have been taken almost verbatim from Araki [**Ara76**].

The relationship between frequency-domain and time-domain inequalities has a long history. The special cases of $(0, I, 0)$ dissipativeness and $(-I, 0, I)$ dissipativeness follow from well-known results for linear passive circuits; see for example Anderson and Vongpanitlerd [**AV73**]. Based on these known results, there was undoubtedly a feeling widespread in the systems theory community that Theorems 25 and 27 were "intuitively obvious"; but a careful treatment took some time to appear. For a detailed discussion, see the optimal control results of J.C. Willems [**Wil71**]. The assumption that leads to condition 8 is from Moylan [**Moy75**], based on a result by Willems [**Wil74**]. Without this assumption, the frequency domain condition becomes more complicated [**And75**].

CHAPTER 9

Simple Nonlinear Systems

1. Introduction

The well-known Popov condition for stability is based on the following scenario. We have a feedback system in which the forward path is a linear system, and the reverse path is a memoryless nonlinearity. There are, as we have seen, well-known sufficient conditions for stability of such systems; but the conditions are not necessary, and sometimes the known tests fail even though the system is stable. Is there some other approach we can try?

One other approach is to augment the system by introducing a first order linear function into the feedback path, and the inverse of that function into the forward path. If this modified system can be shown to be stable, then it is a simple exercise to show that the original system is stable, provided that the augmentation did not create an unstable pole-zero cancellation.

Thus, we are motivated to look for dissipativeness conditions for the combination of a memoryless nonlinearity and a first-order linear system. While doing that, we might as well look more general forms of first-order nonlinear systems.

The obvious systems to study are those with state equations

$$\begin{aligned} \dot{x} &= f(x) + G(x)u \\ y &= h(x) + J(x)u \end{aligned} \tag{9}$$

where x is a scalar. We already have dissipativeness conditions for such systems, in the form of a set of equations that must be satisfied. In the first-order case, there is at least some prospect of finding explicit solutions.

It will turn out that the analysis is complicated for those values of x for which $G(x) = 0$, essentially because control is lost as the state passes through such points. (The analysis is possible; it is actually the statement of results that becomes complicated.) Because that situation is likely to be rare in practice, a standing assumption throughout this chapter is that $G(x) \neq 0$ for all x.

As in the last chapter, we shall proceed by first looking for cyclodissipativeness conditions. If these can be found, we can then proceed to ask whether at least one virtual storage function is a storage function; that is, whether it has nonnegative values for all values of the state.

2. The class of interesting (Q,S,R) triples

It has been mentioned before that not all possible (Q, S, R) triples are interesting. If we consider the function

$$f(u, y) = y^T Q y + 2y^T S u + u^T R u$$

with u and y considered as independent variables — that is, not constrained to be the input and output of a system — then it should be obvious that

- If $f(u, y) \geq 0$ for all u and all y, then *every* system is (Q, S, R) dissipative, regardless of its input-output map;

- If $f(u, y) < 0$ for all u and all y, then *no* system can be (Q, S, R) dissipative.

Thus, the only triples that interest us are those with the property that $f(u, y)$ is negative for some values of u and y, and positive for some other values.

In the general case where Q, S, and R are matrices it is not clear how to express this condition more precisely. In the scalar (single input, single output) case, we can pin it down exactly. Note that

$$f(u, y) = \left(Qz^2 + 2Sz + R\right) u^2 \text{ where } z = y/u$$

from which it is clear that

- If $Q = 0$, $f(u, y)$ can be given both positive and negative values, by choice of u, iff $S \neq 0$;
- If $Q \neq 0$, $f(u, y)$ can be given both positive and negative values iff the equation $Qz^2 + 2Sz + R = 0$ has two distinct real solutions.

It takes only elementary algebra to see that both cases can be covered with a single condition: the triple (Q, S, R) is "interesting" iff $S^2 - QR > 0$. We shall therefore henceforth be making that assumption.

3. General conditions for cyclodissipativeness

For the system 9 to be (Q, S, R) cyclodissipative, there must exist a scalar storage function $\phi(x)$ and vector functions $\ell(x)$ and $W(x)$ such that

$$\begin{aligned} f(x)\frac{d\phi(x)}{dx} &= Qh(x)^2 - \ell(x)^T \ell(x) \\ \frac{1}{2}G(x)\frac{d\phi(x)}{dx} &= h(x)(QJ(x) + S) - \ell(x)^T W(x) \\ QJ(x)^2 + 2SJ(x) + R &= W(x)^T W(x) \end{aligned}$$

We can cast this into a more convenient form by eliminating $\ell(x)$ and $W(x)$, to produce the matrix inequality

$$\begin{bmatrix} Qh(x)^2 - m(x)f(x) & (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) \\ (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) & \hat{R}(x) \end{bmatrix} \geq 0$$

where $m(x) = \frac{d\phi(x)}{dx}$, and $\hat{R}(x) = R + 2SJ(x) + QJ(x)^2$. Clearly one necessary condition for a solution to exist is $\hat{R}(x) \geq 0$ for all x, but beyond this point the conditions are going to depend on whether $\hat{R}(x)$ is zero.

For dissipativeness, we require the further conditions $\phi(0) = 0$ and $\phi(x) \geq 0$ for all x. It is convenient to treat this as a separate issue; that is, to concentrate on the conditions for cyclodissipativeness first, and then to check the sign of the storage function.

To simplify the notation, we shall suppress the explicit x dependence in what follows. That is, we write the above matrix inequality in the form

$$\begin{bmatrix} Qh^2 - mf & (QJ + S)h - \frac{1}{2}Gm \\ (QJ + S)h - \frac{1}{2}Gm & \hat{R} \end{bmatrix} \geq 0 \tag{10}$$

Notice that this is a condition that must be satisfied pointwise for each value of x. In interpreting the meaning of the results, we must of course bear in mind that we are dealing with inequalities which must, in the final summary, be satisfied for all x.

In particular, let us note that it is quite possible for $\hat{R}(x)$ to be zero for some but not all values of x. Since the case where $\hat{R}$ is zero is to be treated as a separate subcase, we must bear in mind the need for the subcases to be unified in the final statement of results.

Before proceeding, let us take note of a useful identity. We have

$$\begin{aligned}\hat{R}Q &= QR + 2QJS + Q^2J^2 \\ &= (QJ+S)^2 - (S^2 - QR) \\ &= \left(QJ + S + \sqrt{S^2 - QR}\right)\left(QJ + S - \sqrt{S^2 - QR}\right)\end{aligned}$$

If either $\hat{R}$ or Q is zero, then $QJ + S = \pm\sqrt{S^2 - QR}$. (Either sign is possible, depending on the parameter values.) Otherwise, we can conclude that the two quantities $QJ + S + \sqrt{S^2 - QR}$ and $QJ + S - \sqrt{S^2 - QR}$ must have the same sign if $Q > 0$, and opposite signs if $Q < 0$. This observation turns out to be useful in a later part of the analysis.

3.1. The case where $\hat{R} = 0$. When $\hat{R}$ is zero, there is a *unique* solution for m, namely

$$m = \frac{2}{G}(QJ + S)h$$

and this solution must satisfy the condition

$$Qh^2 - mf \geq 0$$

or equivalently

$$Qh^2 \geq \frac{2}{G}(QJ + S)hf$$

Let $z = Gh/f$; then the above inequality becomes

$$Qz^2 \geq 2(QJ + S)z$$

Note however that m will become infinite if $G = 0$ and $h \neq 0$, which is one reason we have decided to exclude the case $G = 0$.

3.2. The case where $\hat{R} > 0$. In the more general case $\hat{R} > 0$, the necessary and sufficient condition for cyclodissipativeness is that the determinant of the matrix in inequality (10) be nonnegative. It is easy to show that this determinant can be written as $am^2 + bm + c$, where $a = -\frac{1}{4}G^2$, $b = (QJ + S)Gh - \hat{R}f$, and $c = \hat{R}Qh^2 - (QJ + S)^2h^2 = -(S^2 - QR)h^2$. Observe that both a and c are nonpositive. This means that, if any solution exists for m, then all solutions have the same sign. The condition for a real solution to exist for m is, of course,

$$b^2 - 4ac \geq 0$$

which reduces to

$$\hat{R}f^2 - 2(QJ + S)fGh + Q(Gh)^2 \geq 0$$

If any solution exists, then one such solution is given by

$$m = -\frac{b}{2a} = \frac{2}{G^2}\left((QJ + S)\,Gh - \hat{R}f\right)$$

Observe that this formula remains valid for the case $\hat{R} = 0$. Of course this "solution" is a valid solution only if the discriminant inequality is satisfied. The fact that all solutions have the same sign means that we are not going to lose any generality by focusing on just one particular solution.

As above, let $z = Gh/f$; then the above inequality becomes

$$\hat{R} - 2(QJ + S)z + Qz^2 \geq 0$$

and the form of this inequality is similar to the condition found for the case $\hat{R} = 0$. We can therefore proceed with a single calculation that covers both cases.

3.3. The general case $\hat{R} \geq 0$. In both cases, the necessary and sufficient conditions for cyclodissipativeness are

$$Qz^2 - 2(QJ + S)z + \hat{R} \geq 0$$

where $z = Gh/f$. Depending on the sign of Q, this means that z must lie inside or outside an interval $[b_1, b_2]$. For $Q \neq 0$, the interval bounds are

$$b_1, b_2 = \frac{(QJ+S) \pm \sqrt{(QJ+S)^2 - Q\hat{R}}}{Q} = \frac{(QJ+S) \pm \sqrt{S^2 - QR}}{Q}$$

For $Q = 0$, z must lie in the interval $\left(-\infty, \hat{R}/\left(2S\right)\right]$ if $S > 0$, or in the interval $\left[\hat{R}/\left(2S\right), \infty\right)$ if $S < 0$. We can summarise these results in the following theorem.

THEOREM 47. *Assume that $G(x) \neq 0$ for all x, and that $S^2 > QR$. If $Q = 0$, define $b(x) = J(x) + \frac{1}{2}R/S$, and if $Q \neq 0$ define*

$$b_1(x) = \frac{1}{Q}\left(QJ(x) + S - \sqrt{S^2 - QR}\right)$$
$$b_2(x) = \frac{1}{Q}\left(QJ(x) + S + \sqrt{S^2 - QR}\right)$$

Then the necessary and sufficient conditions for system (9) to be (Q, S, R) cyclodissipative are

(1) $R + 2SJ(x) + QJ(x)^2 \geq 0$ *for all x.*
(2) $z(x) = G(x)h(x)/f(x)$ *lies inside or outside an interval, as follows:*
  (a) *If $Q > 0$ then, for all x, $z(x)$ lies outside the range $(b_1(x), b_2(x))$;*
  (b) *If $Q < 0$ then, for all x, $z(x)$ lies inside the range $[b_2(x), b_1(x)]$;*
  (c) *If $Q = 0$ and $S > 0$ then, for all x, $z(x)$ lies inside the range $(-\infty, b(x)]$;*
  (d) *If $Q = 0$ and $S < 0$ then, for all x, $z(x)$ lies inside the range $[b(x), \infty)$.*

Note that this theorem gives conditions for *cyclodissipativeness*. The conditions imply the existence of a virtual storage function, but we do not yet know anything about its sign.

The possibility $f(x) = 0$ leads to a difficulty in interpreting the above theorem. However, it is easy to show that in that case the required condition is $Qh(x)^2 \geq 0$. If $f(x)$ is zero only for those x for which $h(x)$ is zero then z might well still be well-defined. Otherwise, it should be clear that in the case $f(x) = 0$ the system can be cyclodissipative only if $Q \geq 0$.

4. Conditions for dissipativeness

For the system to be dissipative, we need

$$\phi(x) = \int_0^x m(\sigma)d\sigma \geq 0 \text{ for all } x$$

in addition to the cyclodissipativeness conditions. Although we know that in general the solution for $m(x)$ is not unique, one solution is given by

$$\begin{aligned} m(x) &= \frac{2}{G(x)^2}\left(\left(QJ(x) + S\right)G(x)h(x) - \hat{R}(x)f(x)\right) \\ &= \frac{2}{G(x)^2}\beta(x)f(x) \end{aligned}$$

where

$$\beta(x) = \left(QJ(x) + S\right)z(x) - \hat{R}(x)$$

Consider now the possible values for the sign of $(QJ+S)z-\hat{R}$ when the conditions of theorem 47 are satisfied. Consider first the case $Q \neq 0$. Letting b represent either of the bounds on z, we have

$$\begin{aligned}(QJ+S)b-\hat{R} &= \frac{1}{Q}\left((QJ+S)^2 \pm (QJ+S)\sqrt{S^2-QR}\right)-\hat{R} \\ &= \frac{1}{Q}\left(\hat{R}Q \;+S^2-QR \pm (QJ+S)\sqrt{S^2-QR}\right)-\hat{R} \\ &= \frac{1}{Q}\left(\;\sqrt{S^2-QR} \pm (QJ+S)\right)\sqrt{S^2-QR}\end{aligned}$$

If $Q<0$ then both bounds on β are nonpositive (or strictly negative if $\hat{R}>0$), while if $Q>0$ the two bounds have opposite signs.

In the case $Q=0$ we have a one-sided bound b on z, and we get

$$(QJ+S)\,b-\hat{R} = -\frac{1}{2}\hat{R} \leq 0$$

The conclusion so far is that $\beta(x) \leq 0$ for all x if $Q \leq 0$. (In the case $Q>0$ we cannot conclude anything about the sign of $\beta(x)$.) This calculation has been for just one particular solution for $m(x)$. Recall, however, the following property: for any x, if any solution exists for $m(x)$, then all solutions for $m(x)$ have the same sign. This means that our conclusions about the sign properties of β remain valid for any arbitrary solution.

It is difficult to proceed any further in the general case, because we know nothing about the sign of $f(x)$. We can, however, get stronger results in the case where the open-loop system $\dot{x}=f(x)$ is known to be asymptotically stable. Because we are dealing with a first-order system, the necessary and sufficient condition for asymptotic stability is $xf(x)<0$ for all $x \neq 0$. In this case we can conclude that $xm(x) \geq 0$ for all x, and therefore $\phi(x) \geq 0$ for all x, whenever $Q \leq 0$. That is, cyclodissipativeness implies dissipativeness in this case.

When $Q>0$, a sufficient (but perhaps not necessary) condition for the storage function to be nonnegative is $\beta(x) \leq 0$ for all x. The condition required is

$$(QJ(x)+S)\,z(x) \leq \hat{R}(x)$$

and we must combine this with the cyclodissipativeness condition that $z(x)$ lie outside the range $[b_1(x), b_2(x)]$, where

$$\begin{aligned}b_1(x) &= \frac{1}{Q}\left(QJ(x)+S-\sqrt{S^2-QR}\right) \\ b_2(x) &= \frac{1}{Q}\left(QJ(x)+S+\sqrt{S^2-QR}\right)\end{aligned}$$

Recall that b_1 and b_2 have the same sign. In the case where $QJ+S>0$, the two bounds are positive, and we are adding the extra condition

$$z \leq \frac{\hat{R}}{QJ(x)+S} = \frac{2b_1b_2}{b_1+b_2}$$

In the case $QJ+S<0$ we have the same condition with the inequality reversed. It is easy to see that this new bound lies between b_1 and b_2, the net result being that one of the original bounds is superseded while the other remains in force.

We can summarise these results in the following theorem.

THEOREM 48. *Assume that $G(x) \neq 0$ for all x, and that $S^2>QR$. Suppose also that the free system $\dot{x}=f(x)$ is asymptotically stable. If $Q=0$, define $b(x)=$*

$J(x) + \frac{1}{2}R/S$, and if $Q \neq 0$ define

$$b_1(x) = \frac{1}{Q}\left(QJ(x) + S - \sqrt{S^2 - QR}\right)$$
$$b_2(x) = \frac{1}{Q}\left(QJ(x) + S + \sqrt{S^2 - QR}\right)$$

Then system (9) is (Q, S, R) dissipative if

(1) *$R + 2SJ(x) + QJ(x)^2 \geq 0$ for all x.*
(2) *$z(x) = G(x)h(x)/f(x)$ lies inside an interval, as follows:*
   (a) *If $Q < 0$ then, for all x, $z(x)$ lies inside the range $[b_2(x), b_1(x)]$;*
   (b) *If $Q = 0$ and $S > 0$ then, for all x, $z(x)$ lies inside the range $(-\infty, b(x)]$;*
   (c) *If $Q = 0$ and $S < 0$ then, for all x, $z(x)$ lies inside the range $[b(x), \infty)$;*
   (d) *If $Q > 0$ then $z(x)$ lies inside the range $(-\infty, b_1(x)]$ for all x such that $QJ(x) + S > 0$, and inside the range $[b_2(x), \infty)$ for all x such that $QJ(x) + S < 0$.*

In the case $Q \leq 0$, these conditions are both necessary and sufficient for dissipativeness. For $Q > 0$, the conditions are sufficient but not necessary.

Example 1. *Consider the system*

$$\begin{aligned} \dot{x} &= -x^3 + u \\ y &= \frac{ax^3}{1 + x^2} \end{aligned}$$

From this we can calculate

$$z(x) = \frac{G(x)h(x)}{f(x)} = -\frac{a}{1 + x^2}$$

Obviously z lies in the interval $[-a, 0]$. Comparing this with the conditions of Theorem 48, we can see that this matches the case $Q < 0$, with

$$\begin{aligned} b_1 &= \frac{1}{Q}\left(S - \sqrt{S^2 - QR}\right) = 0 \\ b_2 &= \frac{1}{Q}\left(S + \sqrt{S^2 - QR}\right) = -a \end{aligned}$$

If we arbitrarily choose $S = 1$, then we get $R = 0$ and $Q = -2/a$. The other conditions of the theorem are obviously satisfied, so the system is $(-2/a, 1, 0)$ dissipative. One possible storage function can be calculated via

$$\begin{aligned} m(x) &= \frac{2}{G^2}\left(\left(QJ + S\right)Gh - \hat{R}f\right) = \frac{2ax^3}{1 + x^2} \\ \phi(x) &= \int_0^x \frac{2a\xi^3}{1 + \xi^2} d\xi = ax^2 - a\ln(1 + x^2) \end{aligned}$$

5. Systems with linear dynamics

A case of special interest is where the $u \longmapsto x$ mapping is linear, and all the nonlinearity lies in the readout map. That is, G and J are constant, $h(.)$ is possibly nonlinear, and $f(x) = -\alpha x$ for some constant α. To avoid complicating the analysis, let us consider only the cases where $\alpha \geq 0$.

5.1. Systems with a pole in the left half plane. For the case $\alpha > 0$, the results are immediate from theorems 47 and 48.

THEOREM 49. *Suppose that $\alpha > 0$ and $G \neq 0$, and $S^2 > QR$. Then the necessary and sufficient conditions for the system*

$$\begin{aligned} \dot{x} &= -\alpha x + Gu \\ y &= h(x) + Ju \end{aligned} \tag{11}$$

to be (Q, S, R) cyclodissipative are

(1) $R + 2SJ + QJ^2 \geq 0$.
(2) *$h(x)$ lies inside or outside a sector, as follows:*
  (a) *If $Q > 0$ then $Gh(x)$ lies outside the sector (k_2, k_1);*
  (b) *If $Q < 0$ then $Gh(x)$ lies inside the sector $[k_1, k_2]$;*
  (c) *If $Q = 0$ and $S > 0$ then $Gh(x)$ lies inside the sector $[k, \infty)$;*
  (d) *If $Q = 0$ and $S < 0$ then $Gh(x)$ lies inside the sector $(-\infty, k]$;*

where

$$k_1 = -\frac{\alpha}{Q}\left(QJ + S - \sqrt{S^2 - QR}\right)$$
$$k_2 = -\frac{\alpha}{Q}\left(QJ + S + \sqrt{S^2 - QR}\right)$$

if $Q \neq 0$, and $k = -\alpha\left(J + \frac{1}{2}R/S\right)$ if $Q = 0$.

THEOREM 50. *Suppose that $\alpha > 0$ and $G \neq 0$, and $S^2 > QR$. Then the system*

$$\begin{aligned} \dot{x} &= -\alpha x + Gu \\ y &= h(x) + Ju \end{aligned} \tag{12}$$

is (Q, S, R) dissipative if

(1) $R + 2SJ + QJ^2 \geq 0$.
(2) *$h(x)$ lies inside a sector, as follows:*
  (a) *If $Q < 0$ then $Gh(x)$ lies inside the sector $[k_1, k_2]$;*
  (b) *If $Q = 0$ and $S > 0$ then $Gh(x)$ lies inside the sector $[k, \infty)$;*
  (c) *If $Q = 0$ and $S < 0$ then $Gh(x)$ lies inside the sector $(-\infty, k]$;*
  (d) *If $Q > 0$ and $QJ + S > 0$, then $Gh(x)$ lies inside the sector $[k_1, \infty)$;*
  (e) *If $Q > 0$ and $QJ + S < 0$. then $Gh(x)$ lies inside the sector $(-\infty, k_2]$;*

where

$$k_1 = -\frac{\alpha}{Q}\left(QJ + S - \sqrt{S^2 - QR}\right)$$
$$k_2 = -\frac{\alpha}{Q}\left(QJ + S + \sqrt{S^2 - QR}\right)$$

if $Q \neq 0$, and $k = -\alpha\left(J + \frac{1}{2}R/S\right)$ if $Q = 0$.

We can also say something about the signs of the sector bounds.

- If $Q < 0$ then $k_1 \leq 0 \leq k_2$;
- If $Q = 0$ and $S > 0$ then $k \leq 0$;
- If $Q = 0$ and $S < 0$ then $k \geq 0$;
- If $Q > 0$ and $QJ + S > 0$ then $k_2 \leq k_1 \leq 0$;
- If $Q > 0$ and $QJ + S < 0$ then $k_1 \geq k_2 \geq 0$.

One practical application of these results is where we have a linear system with transfer function $\frac{1}{1+\alpha s}$ followed by a sector nonlinearity h. In this case we can apply the theorems with $G = \alpha$ and $J = 0$. Note that the sector bounds for h are then independent of α — a feature that leads, among other results, to the Popov criterion.

The foregoing results are phrased in such a way that we start with a given (Q, S, R) and then have to derive the sector bounds. Often in practice the question is the other way around: given the sector bounds, find a triple such that the system is (Q, S, R) dissipative. Suppose, then, that we have a linear system with transfer function $\frac{1}{1+\alpha s}$ (with $\alpha > 0$) followed by a memoryless nonlinearity h in the sector $[h_1, h_2]$. From the last theorem we can conclude the following.

- If $h_1 = -\infty$ and $h_2 \geq 0$ then the system is $(Q, -1, (2 - Qh_2)\, h_2)$ dissipative for any Q in the range $0 \leq Q \leq -2/h_1$;
- If $h_1 = -\infty$ and $h_2 < 0$ then the system is $(0, -1, 0)$ dissipative;
- If $h_1 \leq h_2 < 0$ then the system is $(-1, \frac{1}{2}h_1, 0)$ dissipative;
- If $h_1 \leq 0 \leq h_2$ then the system is $(-1, \frac{1}{2}(h_1 + h_2), -h_1 h_2)$ dissipative;
- If $0 < h_1 \leq h_2$ then the system is $(-1, \frac{1}{2}h_2, 0)$ dissipative;
- If $h_1 \leq 0$ and $h_2 = \infty$ then the system is $(Q, 1, -(Qh_1 + 2)\, h_1)$ dissipative for any Q in the range $0 \leq Q \leq -2/h_1$;
- If $h_1 > 0$ and $h_2 = \infty$ then the system is $(0, 1, 0)$ dissipative.

The dissipativeness parameters are never unique. In particular, if a system is (Q, S, R) dissipative then it is also (Q_1, S, R_1) dissipative for any $Q_1 \geq Q$ and $R_1 \geq R$. In the above list we have given only the "least conservative" results, i.e. those corresponding to the smallest possible Q and R. In some cases, where there is a trade-off between Q and R, there is a range of reasonable choices.

5.2. An integrator plus nonlinearity. The following result is for the case $\alpha = 0$.

THEOREM 51. *The necessary and sufficient conditions for the system*

$$\begin{aligned} \dot{x} &= Gu \\ y &= h(x) + Ju \end{aligned}$$

to be (Q, S, R) cyclodissipative are $R + 2SJ + QJ^2 \geq 0$ and $Q \geq 0$.

PROOF. It is not immediately obvious that Theorem 47 is applicable, because z in that theorem is infinite. Let us therefore go back to the more general condition

$$\begin{bmatrix} Qh(x)^2 - m(x)f(x) & (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) \\ (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) & \hat{R}(x) \end{bmatrix} \geq 0$$

where $m(x) = \frac{d\phi(x)}{dx}$ and $\hat{R}(x) = R + 2SJ(x) + QJ(x)^2$. In the present case this simplifies to

$$\begin{bmatrix} Qh(x)^2 & (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) \\ (QJ(x) + S)h(x) - \frac{1}{2}G(x)m(x) & \hat{R}(x) \end{bmatrix} \geq 0$$

Obviously a necessary condition for this to be satisfied is $Q \geq 0$ and $\hat{R} \geq 0$. These conditions are also sufficient, because we are free to choose $m(x)$ such that the off-diagonal elements are zero. □

Let us now consider the question of dissipativeness, as opposed to cyclodissipativeness. If any solution exists for $m(x)$, then one solution is given by

$$m(x) = \frac{2(QJ + S)}{G} h(x)$$

Clearly, a sufficient condition for dissipativeness is that $h(x)$ lie in the sector $[0, \infty)$ or $(-\infty, 0]$, depending on the sign of $(QJ + S)/G$.

THEOREM 52. *The system*

$$\begin{aligned} \dot{x} &= Gu \\ y &= h(x) + Ju \end{aligned}$$

is (Q, S, R) *dissipative provided that* $R + 2SJ + QJ^2 \geq 0$, $Q \geq 0$, *and*

- *if* $(QJ + S)/G > 0$ *then* $h(x)$ *lies in the sector* $[0, \infty)$*;*
- *if* $(QJ + S)/G < 0$ *then* $h(x)$ *lies in the sector* $(-\infty, 0]$.

It is easily verified that the case $QJ + S = 0$ cannot occur. As in an earlier theorem, the sector condition is sufficient but not necessary. A possibility which is not covered by this theorem is illustrated in the following example.

Example 2. *An interesting extreme case is the system*

$$\begin{aligned} \dot{x} &= u \\ y &= h(x) \end{aligned}$$

where $h(x)$ *is defined by*

$$h(x) = \begin{cases} 1 & \text{if} \quad x \in (N, N+1] \quad \text{and } N \text{ is even} \\ -3 & \text{if} \quad x \in (N, N+1] \quad \text{and } N \text{ is odd} \end{cases}$$

This is a case where $h(x)$ *not only fails to satisfy the sector condition, but in fact* $\int_0^x h(\sigma)d\sigma$ *diverges to* $-\infty$ *as* x *increases. Nevertheless, it can be shown that this system is* $(1, 2, 1)$ *dissipative, with storage function*

$$\phi(x) = \begin{cases} 6(x - N) & \text{if} \quad x \in (N, N+1] \quad \text{and } N \text{ is even} \\ 6 - 6(x - N) & \text{if} \quad x \in (N, N+1] \quad \text{and } N \text{ is odd} \end{cases}$$

This example takes advantage of the fact that there is a range of possible solutions for $m(x)$.

CHAPTER 10

Additional results

In this chapter we collect together a number of results that do not really belong anywhere else. We begin with some extensions of the stability results earlier in this book. Next, we look at a relationship between dissipativeness and an inverse problem in optimal control, followed by a treatment of the concept of dissipation delay. Finally, we present a structure result that shows how a passive system can be decomposed as the interconnection of a memoryless passive system and a lossless system.

1. Relaxed stability tests

The results in this section come from [**Moy81**], but they were strongly influenced by related results by Vidyasagar[**Vid79a**].

Let us recall our general stability result for interconnected systems. We have N subsystems and we assume that subsystem i is (Q_i, S_i, R_i) dissipative, for $i = 1..N$. We define an overall input vector u which is a column vector of all the individual subsystem inputs u_i, and similarly for the outputs; and we have a connection equation

$$u = u_e - Hy$$

where u_e represents the external inputs, and H is an $N \times N$ matrix. We then form the matrix

$$\widehat{Q} = Q - SH - H^T S^T + H^T R H$$

and conclude that the overall system is stable if $\widehat{Q} < 0$. Here, "stable" means either input-output stable, or asymptotically stable in the sense of Lyapunov, depending on the system model we are working with.

What if we only have $\widehat{Q} \leq 0$? In this section, we show that that too can sometimes imply stability.

1.1. Combining dissipativeness parameters. It is very common to find that a system is dissipative with respect to two distinct sets of (Q, S, R) parameters. For example, it might happen that a system is both passive and finite gain. Suppose, then, that subsystem i is both $(Q_i^{(1)}, S_i^{(1)}, R_i^{(1)})$ dissipative and $(Q_i^{(2)}, S_i^{(2)}, R_i^{(2)})$ dissipative. We do not, however, require that a second set of dissipativeness parameters be found for every subsystem. For those subsystems for which we cannot find a second set of parameters, we can simply set $(Q_i^{(2)}, S_i^{(2)}, R_i^{(2)}) = (0, 0, 0)$. Obviously, that means that subsystem i is also $(Q_i^{(1)} + \alpha Q_i^{(2)}, S_i^{(1)} + \alpha S_i^{(2)}, R_i^{(1)} + \alpha R_i^{(2)})$ dissipative, for any real scalar $\alpha \geq 0$.

Repeating the $\widehat{Q}$ calculation with these new parameters, we get

$$\begin{aligned}
\widehat{Q} &= Q^{(1)} + \alpha Q^{(2)} - \left(S^{(1)} + \alpha S^{(2)}\right) H - H^T \left(S^{(1)} + \alpha S^{(2)}\right)^T \\
&\quad + H^T \left(R^{(1)} + \alpha R^{(2)}\right) H \\
&= \hat{Q}^{(1)} + \alpha \hat{Q}^{(2)}
\end{aligned}$$

where $\hat{Q}^{(1)}$ is the original $\hat{Q}$, and $\hat{Q}^{(2)}$ is what would have been calculated using only the second set of parameters.

It follows, obviously, that we can conclude stability if there exists some $\alpha \geq 0$ such that $\hat{Q}^{(1)} + \alpha\hat{Q}^{(2)} < 0$.

Typically $\hat{Q}^{(2)}$ will have no special sign properties, but we can always express it as $\hat{Q}^{(2)} = Q_B - Q_C$, where $Q_B \leq 0$ and $Q_C \leq 0$. The choice of Q_B and Q_C will not always be unique, but that turns out not to matter in what follows.

Our key result depends on the following two lemmas. The first of these is from [**Vid79b**], although we prefer a slightly different proof.

LEMMA 11. *Let $A \geq 0$ and $C \geq 0$ be two Hermitian $n \times n$ matrices such that*

$$\textit{rank} \begin{bmatrix} A & C \end{bmatrix} = \textit{rank} \; [A]$$

Then there exists a real $\alpha > 0$ such that $A - \alpha C \geq 0$.

PROOF. It is always possible to find a nonsingular matrix T such that

$$TA = \begin{bmatrix} A_1 \\ 0 \end{bmatrix}$$

where A_1 has full row rank. (If A already has full row rank, then $A_1 = A$ and there are no zero rows.) Let us partition TC in the same way, as

$$TC = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

Now, premultiplication by a nonsingular matrix does not change the rank of a matrix, so the rank condition can be written as

$$\text{rank}\, T \begin{bmatrix} A & C \end{bmatrix} = \text{rank} \; [TA]$$

or

$$\text{rank} \begin{bmatrix} A_1 & C_1 \\ 0 & C_2 \end{bmatrix} = \text{rank} \begin{bmatrix} A_1 \\ 0 \end{bmatrix}$$

and therefore $C_2 = 0$, because A_1 has full row rank.

The premultiplication by T is equivalent to performing a series of elementary row operations on A. If we now perform the same operations on the columns of A, we get

$$TAT^* = \begin{bmatrix} A_{11} & A_{12} \\ 0 & 0 \end{bmatrix}$$

But TAT^* is Hermitian, so $A_{12} = 0$ and $A_{11} > 0$. If we perform the same operations on C, we get

$$T\,(A - \alpha C)\,T^* = \begin{bmatrix} A_{11} - \alpha C_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

Since $A_{11} > 0$ and $C_{11} \geq 0$, we will get $A_{11} - \alpha C_{11} > 0$ for any sufficiently small $\alpha > 0$, and this is sufficient to prove that $A - \alpha C \geq 0$. ☐

A more careful examination of the conditions shows that the range of α for which this works is

$$0 < \alpha < \frac{\lambda_{snz}(A)}{\lambda_{\max}(C)}$$

where $\lambda_{\max}(C)$ is the largest eigenvalue of C, and $\lambda_{snz}(A)$ is the smallest nonzero eigenvalue of A. These are real positive numbers, because non-negative definite Hermitian matrices have real nonnegative eigenvalues.

LEMMA 12. *Let A, B, and C be three symmetric nonnegative definite $n \times n$ matrices such that*

$$\begin{aligned} rank \begin{bmatrix} A & B \end{bmatrix} &= n \\ rank \begin{bmatrix} A & C \end{bmatrix} &= rank\ [A] \end{aligned}$$

Then there exists a real $\alpha \geq 0$ such that $A + \alpha (B - C)$ is positive definite.

PROOF. If $A = 0$ the result is trivially true. Otherwise, Lemma 11 says that we can find an $\alpha > 0$ such that $A - \alpha C \geq 0$. Then for any $y \neq 0$ we can assert that

$$y^* (A - \alpha C)\, y \geq 0$$

and

$$y^* (\alpha B)\, y \geq 0$$

Our rank conditions imply that these two quantities cannot be zero simultaneously, so their sum is always positive. □

The point of Lemma 12 is that we can start with an A that is merely non-negative definite, and with the aid of a minor modification obtain a matrix that is positive definite.

This leads to a stability theorem for the interconnected system. Recall that $\hat{Q}^{(1)}$ was derived from the "primary" dissipativeness parameters, that $\hat{Q}^{(2)}$ comes from a second set of dissipativeness parameters, and that $\hat{Q}^{(2)}$ was decomposed as $\hat{Q}^{(2)} = Q_B - Q_C$, where $Q_B \leq 0$ and $Q_C \leq 0$.

THEOREM 53. *The interconnected system is stable if $\hat{Q}^{(1)} \leq 0$, and if in addition*

$$\begin{aligned} rank \begin{bmatrix} \hat{Q}^{(1)} & Q_B \end{bmatrix} &= n \\ rank \begin{bmatrix} \hat{Q}^{(1)} & Q_C \end{bmatrix} &= rank \begin{bmatrix} \hat{Q}^{(1)} \end{bmatrix} \end{aligned}$$

where n is the sum of the total number of outputs of the subsystems.

PROOF. Obvious from Lemma 12. We trust that the reader is not confused by the change in signs between the lemma and the theorem. □

To apply this in practice, we must choose the supplementary dissipativeness parameters in such a way that Q_B compensates for the rank deficiency of $\hat{Q}^{(1)}$, while not letting Q_C complicate the result. In the most general case it is not obvious how often these conditions will be satisfied, but when we get down to specifics it is not hard to find examples.

For a concrete example, suppose that we have three passive systems and an interconnection matrix

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

The calculation of $\hat{Q}^{(1)}$ is easy in this case:

$$\hat{Q}^{(1)} = -H - H^T = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus $\hat{Q}^{(1)} \leq 0$, but it is singular, so we need a Q_B that will add nonzero entries to the third row and column.

Suppose that the second system has a lower bound on its gain. That is, it satisfies

$$\langle y_2, y_2 \rangle_T \geq \varepsilon \langle u_2, u_2 \rangle_T$$

for some $\varepsilon > 0$, which means that it is $(1/\varepsilon, 0, -1)$ dissipative. This gives a new set of matrices

$$Q^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\varepsilon} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad S^{(2)} = 0 \quad R^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which leads to

$$\widehat{Q}^{(2)} \quad = \quad Q^{(2)} + H^T R H = \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

It can be seen that Q_C adds nothing to the rank of $\widehat{Q}^{(2)}$, while Q_B supplies the desired third row and column, so the conditions of the theorem are satisfied. Note, too, that we do not need to know the value of ε. It is sufficient to know that an $\varepsilon > 0$ exists.

Suppose instead that we knew that the third system had finite gain; that is, that it was $(-1, 0, k^2)$ dissipative. (At this point, you might want to ask yourself why we are putting this condition on the third rather than the second system.) Re-doing the calculation, we get

$$\widehat{Q}^{(2)} \quad = \quad Q^{(2)} + H^T R H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} - k^2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Again, the rank conditions are satisfied, so we can conclude stability.

1.2. Using finite gain constraints. The earliest stability results for interconnected systems tended to be "weak coupling" criteria: the individual subsystems were assumed to be stable in isolation, and then one looked for conditions under which the interconnection did not destroy the stability. We no longer need to restrict ourselves to such conservative approaches. Nevertheless, it is very commonly true that many, perhaps most, of the subsystems in a large-scale system are known to be stable.

Given this, it makes sense to look for results that take advantage of the fact that we know certain subsystems are finite-gain stable, in addition to whatever other dissipativeness properties they have.

Let us suppose, then, that subsystem i is (Q_i, S_i, R_i) dissipative for all i, and that in addition $\|y_i\|_T \leq k_i \|u_i\|_T$ for some but not all i, where the k_i are finite but not necessarily known constants. Let K be a diagonal matrix formed from the k_i, with the diagonal entry set to 0 for those subsystems that are not known to have finite gain. Also let D be a diagonal matrix that keeps track of which subsystems have the finite gain property. We set $D_{ii} = 1$ if output i has the finite gain property, and $D_{ii} = 0$ otherwise. Note that this formulation allows for the case of subsystems with multiple outputs, some of which have the finite gain property and some which do not.

Calculating $\hat{Q}^{(2)}$ as before, we get

$$\hat{Q}^{(2)} = -D + H^T K D K H$$

In addition, we calculate $\hat{Q}^{(1)}$ from the (Q_i, S_i, R_i) parameters as usual. We then have the following corollary of Theorem 53.

THEOREM 54. *With all matrices defined as above, sufficient conditions for stability are*

$$\begin{aligned} \hat{Q}^{(1)} &\leq 0 \\ rank \begin{bmatrix} \hat{Q}^{(1)} & D \end{bmatrix} &= n \\ rank \begin{bmatrix} \hat{Q}^{(1)} \\ DH \end{bmatrix} &= rank \begin{bmatrix} \hat{Q}^{(1)} \end{bmatrix} \end{aligned}$$

PROOF. From Theorem 53, sufficient conditions for stability are $\hat{Q}^{(1)} \leq 0$ and

$$\begin{aligned} \text{rank} \begin{bmatrix} \hat{Q}^{(1)} & D \end{bmatrix} &= n \\ \text{rank} \begin{bmatrix} \hat{Q}^{(1)} & H^T KDKH \end{bmatrix} &= \text{rank} \begin{bmatrix} \hat{Q}^{(1)} \end{bmatrix} \end{aligned}$$

This second condition reduces to the one in the theorem statement because KDK is a diagonal matrix. □

Note that the rank conditions are in this case particularly easy to check. In particular, we do not have to know any of the gain bounds. We only need to know which of the subsystems have the finite gain property, in order to form the D matrix.

2. Connective stability

Connective stability, a concept that appears to have been introduced by Šiljak [**Š78**], refers to a property of an interconnected system where the overall system is stable, and remains stable even if the interconnections are weakened. The precise form of this weakening needs to be defined. Here, we shall assume that, instead of a constant gain H_{ij} between subsystems j and i, we have a nonlinearity in the sector $[-|H_{ij}|, +|H_{ij}|]$. In our analysis, we will allow the nonlinearity to have memory, subject to strict limits on its gain.

As usual, we assume subsystems $y_i = G_i(u_i)$ for $i = 1...N$. To avoid complications, let us assume that each of these subsystems is a single-input single-output system, and that the overall collection of N systems has dissipativeness parameters $(Q_0, S_0, 0)$. (More general cases can be analysed, but the algebra becomes messy.) The interconnection equations are

$$u_i = u_{ei} - a_{ii}y_i - \sum_{j \neq i} a_{ij}\psi_{ij}(y_j)$$

That is, we allow for linear local feedback, but nonlinear interconnections between the subsystems. To model this, we introduce N^2 extra subsystems

$$y_k = \psi_k(u_k) \qquad \text{for } k = N+1...N+N^2$$

These extra subsystems are assumed to satisfy $\|\psi_k(u_k)\|_T \leq \|u_k\|_T$. That is, they have finite gain with a gain bound of 1. Equivalently, they are $(-p_k, 0, p_k)$ dissipative, where we are allowed to choose $p_k > 0$ arbitrarily. Note that N of these new subsystems — the ones that would have represented nonlinear local feedback around each subsystem — are not connected to anything, but we retain them to keep the indexing easier to follow.

Now we have a total of $N + N^2$ subsystems. For the first N subsystems, the new interconnection equations are

$$u_i = u_{ei} - a_{ii}y_i - \sum_{j \neq i} a_{ij}y_{Ni+j}$$

and for the remaining N^2 subsystems, the interconnection equations are

$$u_k = u_{ek} + y_{(k-1) \bmod N+1}$$

where we have introduced extra external inputs u_{ek} to maintain the rule that there must be an external input for each internal input. Putting these equations together, we get

$$u = u_e - Hy = u_e - \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & 0 \end{bmatrix} y$$

where H_{11} is a diagonal matrix, $H_{11} = \mathrm{diag}\{a_{11}, a_{22}, ..., a_{NN}\}$, and H_{21} is a pile of unit matrices:

$$H_{21} = \begin{bmatrix} -I \\ -I \\ \vdots \\ -I \end{bmatrix}$$

The H_{12} submatrix, which has N rows and N^2 columns, is a little more complicated. We have

$$H_{12} = \begin{bmatrix} -A^{(1)} & -A^{(2)} & \dots & -A^{(N)} \end{bmatrix}$$

where

$$A_{ij}^{(k)} = \begin{cases} a_{ij} & \text{if } i = k \text{ and } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

That is, each of the $A^{(k)}$ has nonzero entries only in row k, in such a way that the entire collection of the $A^{(k)}$ forms a stretched-out version of the A matrix.

The overall H matrix is sparse, which simplifies our calculations, but we have to be very careful about keeping track of subscripts.

Now, we are supposing that the original set of N systems is $(Q_0, S_0, 0)$ dissipative. (It is tempting to add an R_0 parameter, but that turns out to complicate the calculations.) The remaining subsystems are $(-P, 0, P)$ dissipative, where P is an arbitrary diagonal positive definite matrix. Thus our overall dissipativeness parameters are

$$\begin{aligned} Q &= \begin{bmatrix} Q_0 & 0 \\ 0 & -P \end{bmatrix} \\ S &= \begin{bmatrix} S_0 & 0 \\ 0 & 0 \end{bmatrix} \\ R &= \begin{bmatrix} 0 & 0 \\ 0 & P \end{bmatrix} \end{aligned}$$

Now we can calculate the value of the matrix

$$\widehat{Q} = Q - SH - H^T S^T + H^T RH$$

The result turns out to be, without too much effort,

$$-\hat{Q} = \begin{bmatrix} -Q_0 - \sum P_k + 2S_0 H_{11} & S_0 H_{12} \\ H_{12}^T S_0 & P \end{bmatrix}$$

where each P_k is a diagonal positive definite matrix. Now we have to find conditions under which $-\hat{Q} > 0$. This job is simplified by observing that Q_0, S_0, H_{11}, and P_k are all diagonal matrices, and even the off-diagonal blocks are reasonably sparse.

The result depends on quasidominance, a property that is defined in the Appendix.

THEOREM 55. *With the above setting, the overall system is stable if* $AS_0 - \frac{1}{2}Q_0$ *is a quasidominant matrix.*

Proof. The quasidominance condition implies that there exists a set of constants $d_j > 0$ such that

$$d_i a_{ii} S_{0i} - \frac{1}{2} d_i Q_{0i} > \sum_{j \neq i} d_j \left| a_{ij} S_{0j} \right|$$

Let us define

$$K_i = d_i a_{ii} S_{0i} - \frac{1}{2} d_i Q_{0i} - \sum_{j \neq i} d_j \left| a_{ij} S_{0j} \right|$$

Also define $d_{Ni+j} = d_{ij}$ for $i = 1..N$; that is, we simply replicate our N constants d_j an additional N^2 times. Now choose

$$(P_k)_j \quad = \quad \begin{cases} \frac{\varepsilon}{d_j} & \text{if } j = k \\ \frac{d_k}{d_j} \left| S_{0k} a_{jk} \right| + \frac{\varepsilon}{d_j} & \text{if } j \neq k \end{cases}$$

where $\varepsilon > 0$ is a constant whose value is still to be chosen. This leads to

$$\sum_k (P_k)_i = \frac{N\varepsilon}{d_i} + \sum_{k \neq i} \frac{d_k}{d_i} \left| S_{0k} a_{ik} \right|$$

Now, observe that, for $k > 0$, row $Nk + j$ of $-\hat{Q}$ has one diagonal element $(P_k)_j$, and if $j \neq k$ then it has exactly one off-diagonal element $S_{0k} a_{jk}$ in column k. (If $j = k$, there are no off-diagonal elements.) Let $r = Nk + j$. By our definition of P, we have

$$d_r \left(-\hat{Q}_{rr} \right) - \sum_{j \neq r} d_j \left| \hat{Q}_{rj} \right| = d_j (P_k)_j - d_k \left| S_{0k} a_{jk} \right| = \varepsilon$$

That takes care of the bottom N^2 rows of $-\hat{Q}$. For the top N rows, we have

$$\begin{aligned} d_i \left(-\hat{Q}_{ii} \right) - \sum_{j \neq i} d_j \left| \hat{Q}_{ij} \right| &= -d_i Q_{0i} - d_i \sum_k (P_k)_i + 2 d_i S_{0i} a_{ii} - \sum_{j \neq i} d_j \left| S_{0i} a_{ij} \right| \\ &= 2K_i + \sum_{j \neq i} d_j \left| S_{0i} a_{ij} \right| - d_i \sum_k (P_k)_i \\ &= 2K_i - N\varepsilon \end{aligned}$$

Clearly we can make this positive by choosing $0 < \varepsilon < \min_i 2K_i/N$. It follows that $-\hat{Q}$ is a quasidominant matrix, and therefore that $\hat{Q} < 0$. □

As an obvious application of this result, consider the set of equations

$$\frac{dx_i}{dt} = -a_{ii} x_i(t) + \sum_{j \neq i} a_{ij} x_j (t - T_{ij})$$

where the T_{ij} are arbitrary time delays. Here each subsystem is an integrator, therefore passive. A time delay is a finite-gain system with a gain bound of 1. The conditions of the theorem are satisfied, and so we can conclude that this system is stable if the matrix A formed from the a_{ij} is quasidominant.

3. An optimal control problem

For our next result, we have to go back to state equations that are linear in the control. Our interest here is primarily in the *inverse* problem of optimal control: under what conditions is a given feedback control law optimal with respect to a given class of performance indices?

Suppose that we have a system

$$\frac{dx}{dt} = f(x) + G(x)u$$

and we want to minimise the performance index

$$J(x(0)) = \lim_{t_f \to \infty} \left\{ n(x(t_f) + \int_0^{t_f} \left(m(x) + u^T Ru\right) dt \right\}$$

where R is a positive definite symmetric matrix, and $m(\cdot)$ and $n(\cdot)$ have the properties $m(0) = n(0) = 0$, $m(x) \geq 0$ for all x, and $n(x) \geq 0$ for all x. We assume that all of these functions have sufficient smoothness to permit the use of the Hamilton-Jacobi-Bellman approach to finding the optimum. We must also assume that the state space is controllable, because without that condition it is not certain that $J(x)$ is finite.

We are going to require that the control be one that stabilises the system. This can be done by adding the explicit condition $x(t_f) = 0$, or by making the final weighting $n(x(t_f)$ sufficiently large. It turns out, in fact, that for the infinite time problem the choice of $n(\cdot)$ is only of marginal importance, in the sense that different choices of $n(\cdot)$ lead to the same solution; but we shall not explore this and related questions because of a desire to focus on one particular point. What we are going to show is that the optimality of the solution can be expressed as a dissipativeness condition.

It is standard to start with the solution of the corresponding finite-time problem, and then take the limit. Let the optimal performance index be $V(x, t; t_f)$; then the optimal control is found by minimising

$$H(V, x, u) = m(x) + u^T Ru + \nabla V^T f(x) + \nabla V^T G(x)u$$

giving $u = -k(x)$, where

$$k(x) = \frac{1}{2} R^{-1} G(x)^T \nabla V$$

To find V, we must solve the equation

$$\frac{\partial V}{\partial t} + H(V, x, -k(x)) = 0$$

with boundary condition $V(x, t_f; t_f) = n(x(t_f))$. The differential equation simplifies down to

$$\frac{\partial V}{\partial t} + \nabla V^T f(x) - \frac{1}{4} \nabla V^T G(x) R^{-1} G(x)^T \nabla V + m(x) = 0$$

This is for the finite-time problem. Now, it is easy to show that $V(x, t; t_f)$ monotonically approaches a limit $\phi(x)$ as $t_f \to \infty$, that this limit satisfies the same equation but with $\frac{\partial \phi}{\partial t} = 0$, and that the optimal control for the infinite-time problem is still given by the same formula, with ϕ replacing V.

Once we proceed to the infinite-time case, the final weighting function $n(\cdot)$ disappears from the equations. What is happening here is that the limiting equation

$$\nabla \phi(x)^T f(x) - \frac{1}{4} \nabla \phi(x)^T G(x) R^{-1} G(x)^T \nabla \phi(x) + m(x) = 0$$

has multiple solutions. The boundary conditions for the finite-time version of the problem force just one of those solutions to be chosen. It turns out that, of all the solutions, there is only one that produces a stabilising control.

Using the known expression for $k(x)$, the differential equation can also be written as

$$\nabla \phi^T f(x) - k(x)^T Rk(x) + m(x) = 0$$

Now, for any input u, not necessarily optimal,

$$\frac{d\phi}{dt} = \nabla \phi^T f(x) + \nabla \phi^T G(x)u$$

so that the last equation reduces to

$$m(x) + u^T Ru = (u + k(x))^T R (u + k(x)) - \frac{d\phi}{dt}$$

Then for any $T > 0$ we have

$$\int_0^T \left(m(x) + u^T Ru\right) dt = \int_0^T (u + k(x))^T R (u + k(x))\, dt + \phi(x(0)) - \phi(x(T))$$

As a side issue, let us look at the stability of the closed loop system. When u is set equal to the optimal control, the last equation reduces to

$$\phi(x(T)) = \phi(x(0)) - \int_0^T \left(m(x) + u^T Ru\right) dt$$

which says that $\phi(x(t))$ is non-increasing with time. Because $\phi(x) \geq 0$, $\phi(x(t))$ must converge to some finite limit, which means that $x(t)$ converges to a manifold where $\phi(x)$ is constant. For initial states on that manifold, the optimal trajectories are such that $u(t) = 0$ and $m(x(t)) = 0$. With the obvious observability condition, that manifold can only be the origin of the state space, so we have an asymptotically stable system. In this case, the final weighting $n(x(t_f))$ is of so little importance that it might as well be set to zero. Without the observability condition, however, we have to set the terminal weighting to a suitably high value in order to force the optimal solution to be a stabilising solution.

Returning now to the case where u is not necessarily optimal, the fact that $m(x) \geq 0$ for all x means that

$$\phi(x(0)) + \int_0^T \left(k(x)^T Rk(x) + 2k(x)^T Ru\right) dt \geq \phi(x(T))$$

That is, the system with input u and output $k(x)$ is $(R, R, 0)$ dissipative.

Note that this is a result for an open-loop system. We can also express the result in terms of the closed-loop variables. With the loop closed, we have

$$u = u_e - k(x)$$

where u_e is the external input. Now the last inequality becomes

$$\phi(x(0)) + \int_0^T \left(-k(x)^T Rk(x) + 2k(x)^T Ru_e\right) dt \geq \phi(x(T))$$

which means that the optimal closed-loop system, with input u_e and output $k(x)$, is $(-R, R, 0)$ dissipative. That in turn implies that the closed loop has a good margin of stability.

There is at least one more interesting way of arranging this inequality. If we eliminate $k(x)$, we get

$$\phi(x(0)) + \int_0^T \left(u_e^T Ru_e - u^T Ru\right) dt \quad \geq \quad \phi(x(T))$$

In particular, when $x(0) = 0$, this implies

$$\int_0^T u^T Ru\ dt \leq \int_0^T u_e^T Ru_e\ dt$$

What this means is that, for any external input u_e, the feedback is always such as to reduce the input to the original system. Loosely speaking, the feedback is always negative. This is a property of optimal solutions to this class of problem that is not, of course, shared by feedback controls in general.

Optimal feedback of the class just discussed gives systems a number of desirable properties, but only when $m(x) \geq 0$ for all x. It is therefore of interest to look at the inverse problem: given a proposed feedback law $u = -k(x)$, under what conditions does there exist an $m(x) \geq 0$ such that this feedback law is optimal for a performance index of the type we are discussing?

THEOREM 56. *Let R be a positive definite symmetric matrix. Then for the system*

$$\frac{dx}{dt} = f(x) + G(x)u$$

the stabilising control $u = -k(x)$ is an optimal control for a performance index of the form

$$J(x(0)) = \lim_{t_f \to \infty} \left\{ n(x(t_f) + \int_0^{t_f} \left(m(x) + u^T R u\right) dt \right\}$$

for some $m(\cdot)$ and $n(\cdot)$ with the properties $m(0) = n(0) = 0$, $m(x) \geq 0$ for all x, and $n(x) \geq 0$ for all x, if and only if the system

$$\begin{aligned} \frac{dx}{dt} &= f(x) + G(x)u \\ y &= k(x) \end{aligned}$$

is $(R, R, 0)$ dissipative.

PROOF. Necessity has already been shown. For the converse, suppose that we have a $\phi(\cdot)$ satisfying

$$\phi(x(0)) + \int_0^T \left(k(x)^T R k(x) + 2k(x)^T R u\right) dt \geq \phi(x(T))$$

As was done in an earlier chapter, we can turn the inequality into an equality by writing

$$\begin{aligned} \phi(x(0)) \quad &+ \quad \int_0^T \left(k(x)^T R k(x) + 2k(x)^T R u\right) dt \\ &= \quad \phi(x(T)) + \int_0^T \left(\ell(x) + Wu\right)^T \left(\ell(x) + Wu\right) dt \end{aligned}$$

If we reduce this to differential form we discover that W must be zero, so that

$$\begin{aligned} \phi(x(0)) \quad &+ \quad \int_0^T \left(u + k(x)\right)^T R \left(u + k(x)\right) dt \\ &= \quad \phi(x(T)) + \int_0^T \left(\ell(x)^T \ell(x) dt + u^T R u\right) dt \end{aligned}$$

Define $m(x) = \ell(x)^T \ell(x)$ and $n(x) = \phi(x)$. Then

$$\begin{aligned} \phi(x(0)) \quad &+ \quad \int_0^T \left(u + k(x)\right)^T R \left(u + k(x)\right) dt \\ &= \quad \phi(x(T)) + \int_0^T \left(m(x) + u^T R u\right) dt \end{aligned}$$

If $u = -k(x)$, this reduces to

$$\phi(x(0)) = \phi(x(T)) + \int_0^T \left(m(x) + u^T R u\right) dt$$

which becomes, in the limit as $T \to \infty$,

$$\phi(x(0)) = \int_0^\infty \left(m(x) + u^T R u\right) dt$$

This certainly makes it plausible that $u = -k(x)$ is an optimal control that minimises this last integral. To be certain of this, suppose that there is some other stabilising control u_1 such that the resulting trajectory leads to

$$\int_0^\infty \left(m(x) + u_1^T R u_1\right) dt < \phi(x(0))$$

With this control, we have

$$\phi(x(0)) + \int_0^\infty (u_1 + k(x))^T R (u_1 + k(x))\, dt < \phi(x(0))$$

which is impossible because $R > 0$. □

The above result was proved in [**MA73**], although without using the language of dissipativeness. By stating the result in terms of dissipativeness, we have obtained a much simpler proof.

4. Dissipation delay

It was noted in Chapter 3 that a dissipative system can have (and usually does have) multiple storage functions. One way to interpret this fact is to suppose that the different storage functions reflect different storage mechanisms, and therefore different internal realisations of a given input-output relationship. This is consistent with the idea that dissipativeness is an input-output property, while a storage function is a function of the internal state.

Let us define a dissipation function $D(x, u, t_0, t_1)$ via the "conservation of energy" equation

$$\phi(x(t_0)) + \int_{t_0}^{t_1} w(t)dt = \phi(x(t_1)) + D(x(t_0), u, t_0, t_1) \tag{13}$$

where of course

$$w(t) = y^T Q y + 2y^T S u + u^T R u$$

Obviously dissipativeness implies that $D(x, u, t_0, t_1) \geq 0$ for all x and u, and all $t_1 \geq t_0$. In fact this is also true for a cyclodissipative system, although in that case the motivation for defining a dissipation function is less strong.

If $[\phi, D]$ is any pair satisfying 13, we call $[\phi, D]$ a realisation of the system. To narrow this down a little further, we can note that the integral in equation 13 depends only on the input and output, and not on the internal state.

LEMMA 13. *If $[\phi_1, D_1]$ and $[\phi_2, D_2]$ are any two realisations of the same dissipative system, then*

$$D_1(x_0, u, t_0, t_1) = D_2(x_0, u, t_0, t_1)$$

for any x_0 for which $\phi_1(x_0)$ and $\phi_2(x_0)$ are both finite, any $t_1 \geq t_0$, and any u such that $x(t_1) = x(t_0) = x_0$.

PROOF. If $x(t_1) = x(t_0)$, then

$$D_2(x(t_0), u, t_0, t_1) = \int_{t_0}^{t_1} w(t)dt = D_1(x(t_0), u, t_0, t_1)$$

□

This says that the dissipated energy, for cyclic motions only, depends only on the input-output relationship and not on the internal state. It might happen, though, that some realisations dissipate most of the energy in the early part of the cycle, while others initially store most of the input energy and dissipate it only later. This motivates the following definition.

DEFINITION 21. *Let* $[\phi_1, D_1]$ *and* $[\phi_2, D_2]$ *be two realisations of the same dissipative system. Then* $[\phi_1, D_1]$ *has less dissipation delay than* $[\phi_2, D_2]$*, written* $[\phi_1, D_1] \prec [\phi_2, D_2]$*, if*

$$D_1(0, u, t_0, t_1) \geq D_2(0, u, t_0, t_1)$$

for all u *and all* $t_1 \geq t_0$*.*

That is, less dissipation delay means that energy is dissipated sooner in time. Of course this is only a partial ordering; it is possible that the two realisations cannot be compared in this way.

Where a definite ordering is possible, we have the following result.

THEOREM 57. *If* $[\phi_1, D_1]$ *and* $[\phi_2, D_2]$ *are two realisations of the same dissipative system, then* $[\phi_1, D_1] \prec [\phi_2, D_2]$ *iff* $\phi_1(x) \leq \phi_2(x)$ *for all* x*.*

PROOF. This follows directly from equation 13. □

This makes sense intuitively if we interpret $\phi_1(x) \leq \phi_2(x)$ as meaning that $[\phi_1, D_1]$ has a smaller storage capacity than $[\phi_2, D_2]$. Thus, the first realisation has more of a tendency to dissipate energy as soon as it arrives, rather than storing it for later.

As an example, consider the first order system with state equations

$$\begin{aligned} \frac{dx}{dt} &= -x + u \\ y &= x + \frac{1}{2}u \end{aligned}$$

This system is $(0, \frac{1}{2}, 0)$ dissipative. That is, it is passive. Applying the criteria of Chapter 3, we find that the available storage is $\phi_a(x) = \frac{(2-\sqrt{3})}{2}x^2$ and the required supply is $\phi_r(x) = \frac{(2+\sqrt{3})}{2}x^2$. Thus, every quadratic storage function has the form $\frac{1}{2}Cx^2$, where $(2-\sqrt{3}) \leq C \leq (2+\sqrt{3})$. (There might also be non-quadratic storage functions, but if so these are more difficult to find.) The dissipation function is then

$$\begin{aligned} D_1(x_0, u, 0, T) &= \int_0^T \left[Cx^2 + (1-C)ux + \frac{1}{2}u^2\right] dt \\ &= \int_0^T \left[C\,(x - \gamma u)^2 + Ru^2\right] dt \end{aligned}$$

where $\gamma = \frac{C-1}{2C}$ and $R = \frac{1}{2} - C\gamma^2$. Note that $R \geq 0$ iff $(2-\sqrt{3}) \leq C \leq (2+\sqrt{3})$, and $R = 0$ at the two extremes of the range. For values of C outside this range, it is possible to make the "dissipation" negative. That would correspond to finding a realisation that was externally dissipative but internally non-dissipative.

For a given value of C, it is a meaningful exercise to find that u that will minimise this expression for the dissipation. For small C, the low-dissipation trajectories turn out to be those for which $\|x\|$ is decreasing, and for large C the opposite is true. In the extreme case $C = (2+\sqrt{3})$, where the storage function is the required supply, it turns out to be possible to drive the state from the origin to any desired final state with an arbitrarily small amount of dissipation, although driving the state back to the origin creates a non-negligible amount of dissipation. At the opposite extreme, the storage function is the available energy. In that case we must expend a non-negligible amount of energy to get from the origin to a desired state, but from there we can extract all of the stored energy with negligible dissipation.

To provide a physical example, suppose that the system is an electrical circuit, where u is the input current and y is the voltage at the terminals. A realisation

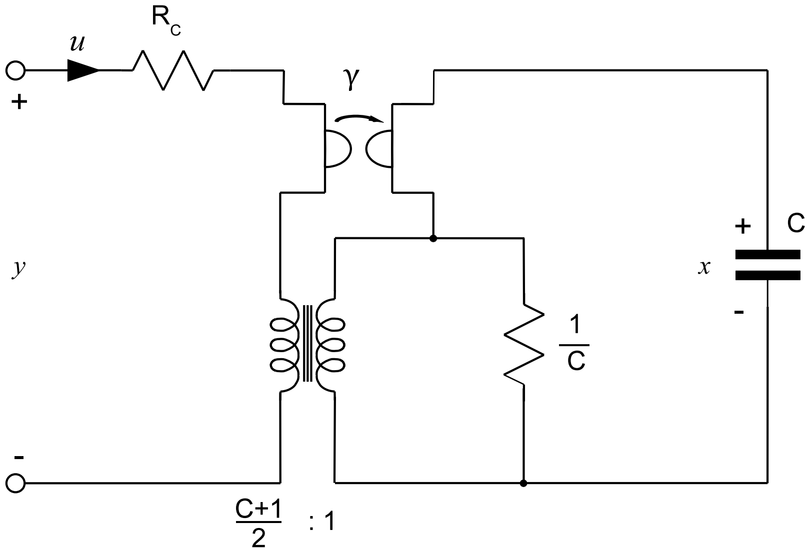

FIGURE 1. Realisation of a passive circuit

where the stored energy is $\frac{1}{2}Cx^2$ can be found by letting x be a capacitor voltage, and then expressing x as the sum of two voltage drops. The result is shown in Figure 1. The power dissipated in the two resistors is obviously Ru^2 and $C\left(x - \gamma u\right)^2$ respectively, which agrees with the expression for $D_1(x_0, u, 0, T)$.

Note that this realisation uses a gyrator, a non-reciprocal component. In the special case $C = 1$ the gyrator disappears, and we have a much simpler internally reciprocal circuit. In this book we have not paid any attention to the relationship between internal and external reciprocity, but this subject is covered — although only for linear systems — in the treatment by Willems [**Wil72**]. The extension of the concept of reciprocity to nonlinear systems is still a poorly understood problem.

For linear systems, we can give a transfer function interpretation to the concept of dissipation delay. For our example, we can define two "dissipation outputs"

$$\begin{aligned} y_1 &= \sqrt{R}u \\ y_2 &= \sqrt{C}(x - \gamma u) \end{aligned}$$

In our electrical circuit, these are normalised versions of the two resistor currents. Let $Y_1(s)$, $Y_2(s)$, and $U(s)$ be the Laplace transforms of the corresponding time-domain variables. Then

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} \sqrt{R} \\ \sqrt{C}\frac{(1-\gamma)-\gamma s}{1+s} \end{bmatrix} U(s) = V(s)U(s)$$

As C varies from $(2-\sqrt{3})$ to $(2+\sqrt{3})$, γ varies monotonically from $-\frac{1}{2}(\sqrt{3}+1)$ to $+\frac{1}{2}(\sqrt{3}-1)$. As a result, the phase lag of $Y_2(s)$ with respect to $U(s)$ increases monotonically with C. This gives us a connection between dissipation delay and phase lag. But we also have

$$V(-s)^T V(s) = \frac{1}{2} + \frac{1}{1-s^2}$$

which does not depend on C. This independence is what we should expect from Lemma 13. The net dissipation rate for a periodic motion is the same for all realisations. Only the phase delay differs from one realisation to another.

The extension of these ideas from one example to the general linear case is left as an interesting — and rather challenging — exercise for the reader.

5. A structure result

A well-known result in linear circuit theory is that a passive circuit can be decomposed as an interconnection of a lossless part (the inductors and capacitors) and a memoryless part. In this section we are going to demonstrate a similar decomposition for an arbitrary passive system. This extends a result of [**AM75**] for the case of state equations that were linear in the control. In hindsight, it appears that that restriction was unnecessary.

Suppose that we have a system with state equations

$$\begin{aligned} \frac{dx}{dt} &= F(x(t), u(t)) \\ y(t) &= H(x(t), u(t)) \end{aligned}$$

Passivity is the same as $(0, \frac{1}{2}I, 0)$ dissipativeness; we have included the factor $\frac{1}{2}$ to make our results more consistent with circuit theory. Let us assume that we have a differentiable storage function. The differential form of the dissipation inequality is then

$$H^T u - \nabla\phi^T F \geq 0$$

To get our result, we need to assume that the map $x \to \nabla\phi(x)$ is invertible. That is, that there exist functions $A(\cdot)$ and $B(\cdot)$ such that $A(\nabla\phi(x)) = x$ and $\nabla\phi(B(x))) = x$. (Under these conditions, it is then easy to show that $A(\cdot)$ and $B(\cdot)$ are the same function.) For linear systems, the invertibility follows from requiring that the state space representation be minimal. For nonlinear systems, we require a case-by-case analysis. We conjecture that convexity of the storage function is a sufficient condition for invertibility of the gradient.

Let us define a subsystem

$$\begin{aligned} \frac{dx}{dt} &= u_1 \\ y_1 &= \nabla\phi(x) \end{aligned}$$

This subsystem is $(0, \frac{1}{2}I, 0)$ lossless. To see why, observe that

$$\int_{t_0}^{t_1} u_1^T(t) y_1(t) dt = \int_{t_0}^{t_1} \frac{d}{dt}\phi(x(t)) dt = \phi(x(t_1)) - \phi(x(t_0))$$

Now, let us define a second memoryless subsystem, with input $\begin{bmatrix} u \\ u_2 \end{bmatrix}$ and output $\begin{bmatrix} y \\ y_2 \end{bmatrix}$, defined by the equations

$$\begin{bmatrix} y \\ y_2 \end{bmatrix} = \begin{bmatrix} H(A(u_2), u) \\ -F(A(u_2), u) \end{bmatrix}$$

Next, let us interconnect the two subsystems via the equations

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Observe that

$$\begin{bmatrix} y_1^T & y_2^T \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

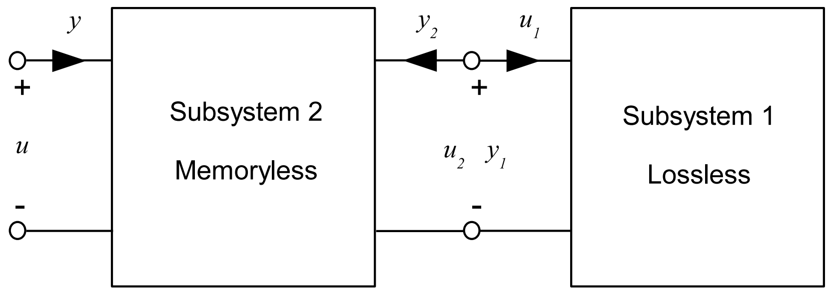

FIGURE 2. Structural decomposition of a passive system

That means that this is a *neutral interconnection*, in the sense that it is passive, memoryless, and lossless. In control theory terms, it is a simple feedback connection. In circuit theory terms, it is a direct connection between the first subsystem and the "2" port of the second subsystem, in such a way that the voltages are equal and the current out of one subsystem is equal to the current into the other.

Putting these equations together, we get an overall result

$$\begin{aligned}\frac{dx}{dt} &= -y_2 = F(x,u)\\ y &= H(x,u)\end{aligned}$$

which confirms that the interconnection of the two subsystems is a realisation of the original state equations. To complete the analysis, observe that

$$\begin{bmatrix} y & y_2 \end{bmatrix}\begin{bmatrix} u \\ u_2 \end{bmatrix} = y^T u + y_2^T y_1 = H^T u - \nabla\phi(x)^T F \geq 0$$

That means that our second subsystem is passive.

The conclusion is that our original passive system can be realised as a neutral connection of a lossless subsystem and a memoryless passive subsystem.

To get a circuit theory interpretation of this result, observe that the interconnection equation describes a cascade connection of subsystem 2 (the memoryless subsystem) and subsystem 1 (the lossless subsystem). The interconnection equations $u_2 = y_1$ and $u_1 = -y_2$ suggest that u_2 and y_1 are voltages and u_1 and y_2 are currents. This in turn suggests that subsystem 1 is a capacitor bank. That, however, is just one possible interpretation. It is equally possible to label the variables in such a way that the lossless subsystem is an inductor bank. An interesting, and so far unsolved, problem would be to allocate the voltage/current port variables in such a way that each subsystem is a reciprocal circuit.

The result in this section is for passive systems only. For more general dissipative systems, we cannot get a comparable result until we can devise a suitable definition for the concept of a "neutral interconnection". This is still an unsolved problem.

APPENDIX A

Some useful matrix results

In this appendix we look at some special classes of matrix, classes that are important in the analysis of interconnected systems. We want to look at two special classes of matrices: M-matrices, and quasidominant matrices. They both have a special "diagonal dominance" property.

In this appendix (although not in the rest of this book) we are particularly concerned with matrices and vectors whose entries have specified sign patterns. We therefore need the following notation. For a row or column vector x, the condition $x > 0$ means that all elements of x are real and positive, and $x \geq 0$ means that all elements of x are real and nonnegative. Similarly, $x < 0$ means $-x > 0$, and $x \leq 0$ means $-x \geq 0$. For a square matrix, $A > 0$ instead means that A is symmetric and positive definite. The determinant of a matrix will be written as $\det A$, while $|A|$ will mean the matrix obtained from A by replacing each element by its absolute value. A *signature matrix* is a diagonal matrix whose diagonal elements are either $+1$ or -1.

A *row permutation* of a matrix A is a reordering of its rows. A simple row permutation, where exactly two rows are swapped, can be obtained by premultiplying by a permutation matrix

$$P = \begin{bmatrix} I & \vdots & \ddots & \vdots & 0 \\ \cdots & 0 & \ddots & 1 & \cdots \\ \ddots & \vdots & I & \vdots & \ddots \\ \cdots & 1 & \cdots & 0 & \cdots \\ 0 & \vdots & \ddots & \vdots & I \end{bmatrix}$$

Any other row permutation can be achieved with a product of such matrices. A permutation matrix has the property $P^{-1} = P$, and the determinant of a permutation matrix is either $+1$ or -1.

A column permutation is produced in the same way, except that we postmultiply by the permutation matrix.

A simple permutation of a square matrix multiplies its determinant by -1. An even number of simple permutations leaves the determinant unchanged. In this appendix we will be only interested in the sort of permutation expressed by PAP, where the columns are permuted in exactly the same way as the rows. This, being a net even number of simple permutations, leaves the determinant unchanged.

Let a set J be some non-empty subset of the index set $\{1, 2, ...N\}$, where N is the number of rows (and columns) of the square matrix A under consideration. Then we define $A(J)$ to be the smaller matrix obtained from A by deleting row i and column i, for every $i \notin J$. The *principal submatrices* of A are, by definition, the collection of all the $A(J)$, where J ranges over every non-empty subset of the index set. (This includes A itself.) The *principal minors* of A are the determinants of all the principal submatrices of A.

For a permutation matrix P, it is easy to see that $(PAP)(J)$ is equal to $P(J)A(J)P(J)$, and that $P(J)$ is itself a permutation matrix. From this it follows that the principal minors of PAP are identical to those of A, although probably enumerated in a different order.

1. A duality result

Let us temporarily move away from square matrices to look at an existence theorem. The relevance of this to matrices with positive principal minors, and thus to stability criteria, will not be immediately clear, but the main result of this section will become important in the following section.

In this section, we will be interested in real matrices, not necessarily square, which have either of the following properties:

- There exists $x > 0$ such that $Ax > 0$.
- There does not exist any $y \geq 0$ with $y \neq 0$ such that $A^T y \leq 0$.

It will turn out that these two conditions are equivalent. For the sake of clarity, we shall break the proof of this into two parts.

Lemma 14. *If matrix A has the property that there exists some $x > 0$ such that $Ax > 0$, then there does not exist any $y \geq 0$ with $y \neq 0$ such that $A^T y \leq 0$.*

Proof. Suppose that there exists an $x > 0$ such that $Ax > 0$, and also a nonzero $y \geq 0$ such that $A^T y \leq 0$. The fact that $Ax > 0$ and $y \geq 0$ means that $x^T A^T y \geq 0$, with equality only if $y = 0$. On the other hand, $x > 0$ and $A^T y \leq 0$ implies $x^T A^T y \leq 0$. These two conflicting inequalities together imply that $y = 0$. □

Lemma 15. *If there does not exist any nonzero $y \geq 0$ such that $A^T y \leq 0$, then there exists some $x > 0$ such that $Ax > 0$.*

Proof. For the sake of considering all possible cases, suppose first that $A = 0$. In that case $A^T y \leq 0$ for any nonzero $y \geq 0$, and $Ax > 0$ can never be satisfied, so the statement of the lemma is vacuously true.

Consider next the case where A has a single row; that is, where $A = b^T$ for some nonzero column vector b. In this case y is a scalar, $A^T y = by$ is a vector, and so the condition $A^T y \leq 0$ for nonzero $y \geq 0$ reduces to saying that all entries of b are nonpositive. The converse, that there does not exist any nonzero $y \geq 0$ such that $A^T y \leq 0$, is the condition that there exists at least one component of b, say b_k, which is positive. In that case, $Ax = b^T x$ can be made positive by choosing x_k large and the other components of x small but positive.

Now we can proceed by induction on the number of rows of A. Let

$$A = \begin{bmatrix} A_1 \\ b^T \end{bmatrix}$$

where b is a column vector. Then

$$A^T y = \begin{bmatrix} A_1^T & b \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = A_1^T y_1 + b y_2$$

where y_2 is a scalar, and

$$Ax = \begin{bmatrix} A_1 x \\ b^T x \end{bmatrix}$$

If there does not exist any $x > 0$ such that $Ax > 0$, then there are two possibilities: either there does not exist any $x > 0$ such that $A_1 x > 0$, or there does not exist any $x > 0$ such that $b^T x > 0$. (Or both, of course.) Let us take the simplest case first. If there is no $x > 0$ such that $b^T x > 0$, then $b \leq 0$. In that case we may choose $y_1 = 0$, $y_2 > 0$ as a demonstration that there exists a nonzero $y \geq 0$ such that $A^T y \leq 0$, which proves our result.

In the remaining case the sign of by_2 is uncertain, but this does not matter. By the inductive hypothesis, the failure of a suitable x to exist implies that there exists some nonzero $y_1 \geq 0$ such that $A_1^T y_1 \leq 0$. We may then set $y_2 = 0$ to complete the inductive step. □

Putting these two results together, we have our duality result.

THEOREM 58. *For a real, not necessarily square, matrix A, the following two statements are equivalent.*

- There exists $x > 0$ such that $Ax > 0$.
- There does not exist any $y \geq 0$ with $y \neq 0$ such that $A^T y \leq 0$.

PROOF. This is simply a restatement of Lemmas 14 and 15. □

Duality results similar to this have been around for a long time. In fact, it turns out that Lemma 15 can also be derived from an early result by Stiemke [**Sti15**] on the existence of solutions to linear homogenous equations.

2. Matrices with positive principal minors

There has been considerable interest over the years in conditions that will guarantee that all principal minors of a real square matrix are positive. For a symmetric matrix the answer is well-known: the matrix has positive principal minors iff the matrix is positive definite. For not necessarily symmetric matrices the answer is less simple. A number of ways of looking at the question were presented by Fiedler and Pták [**FP62**]. Here we want to look at one extra approach which is particularly useful as a lead-in to the later results in this Appendix.

In some of the literature, a real square matrix, all of whose principal minors are positive, is called a P-matrix. It is convenient to adopt that terminology here.

Our first result is by Gale and Nikaidô [**GN65**].

LEMMA 16. *If A is a P-matrix, then the inequalities $Ax \leq 0$, $x \geq 0$ have only the trivial solution $x = 0$.*

PROOF. The result is obvious for a 1×1 matrix, so let us proceed by induction. If A is a P-matrix, then the diagonal entries of A^{-1} are positive. (This is a direct consequence of Cramer's rule for inverting a matrix.) That means that every column of A^{-1} has at least one positive entry. Let b be the first column of A^{-1}, and suppose that we have an $x \geq 0$ such that $Ax \leq 0$. Define α to be the minimum value of $\frac{x_i}{b_i}$, where the minimum is taken over those i for which $b_i > 0$, and let k be the value of i for which the minimum is attained. (Note that $\alpha \geq 0$.) Then we have $x_i \geq \alpha b_i$ for all i, including those i for which $b_i \leq 0$. That is, $z \geq 0$, where $z = x - \alpha b$.

Now observe that, because b is the first column of A^{-1}, $Ab = e_1$, where e_1 is the unit vector

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

From this it follows that $Az = Ax - \alpha Ab = Ax - \alpha e_1 \leq 0$.

Observe that, because z_k is zero, column k of A has no effect on the calculation of Az. Define $\widehat{A}$ to be the principal submatrix of A formed by deleting row and column k, and define $\widehat{z}$ to be z with row k deleted. Then clearly $\widehat{z} \geq 0$ and $\widehat{A}\widehat{z} \leq 0$. From the inductive hypothesis, and because $\widehat{A}$ is also a P-matrix, it follows that

$\widehat{z} = 0$ and therefore $z = 0$. Then $Ax = Az + \alpha e_1 = \alpha e_1$, from which it follows that $Ax \geq 0$. However, we have already asserted that $Ax \leq 0$, so the only possibility is $Ax = 0$. Since A is nonsingular, we must have $x = 0$. □

This leads to a very interesting fact.

THEOREM 59. *If A is a P-matrix, then there exists an $x > 0$ such that $Ax > 0$.*

PROOF. This is a combination of the results of Lemma 16 and Theorem 58. □

The above theorem appears to be widely known, but one rarely sees a proof. It appears that Gale and Nikaidô [**GN65**] were the first to come up with a proof.

We are now in a position to move to an "if and only if" characterisation. The following theorem is taken from [**Moy77**].

THEOREM 60. *A real square matrix A is a P-matrix iff, for every signature matrix S, there exists a vector $x > 0$ such that $SASx > 0$.*

PROOF. If A is a P-matrix, Theorem 59 shows that there exists an $x > 0$ such that $Ax > 0$. Since the principal minors of SAS are identical with those of A, this proves half of the theorem.

For the converse, suppose that for each S there exists an $x > 0$ such that $SASx > 0$. Note that $(SAS)_{ij} = \sigma_i \sigma_j A_{ij}$, where $\sigma_i = \pm 1$ for each i. If we arbitrarily choose $\sigma_i = +1$, the condition implies

$$A_{ii} x_i + \sum_{j \neq i} \sigma_j A_{ij} x_j > 0$$

This must hold for any combination of the σ_j, although the x is allowed to be different for different signature matrices. Let us pick the "worst case" value $\sigma_j = -sgn A_{ij}$, giving

$$A_{ii} x_i > \sum_{j \neq i} |A_{ij}| \, x_j$$

(If $A_{ij} = 0$, the choice of σ_j is irrelevant.) This tells us that $A_{ii} > 0$. That is, all 1×1 principal minors are positive. We can now complete the proof by induction. Let us suppose that it is known that all $m \times m$ principal minors are positive. Partition the matrix A as

$$A = \begin{array}{c} \left[\begin{array}{ccc} A_{11} & A_{12} & a_{13} \\ A_{21} & A_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right] \end{array} \begin{array}{l} \\ \updownarrow m \\ \updownarrow 1 \end{array}$$

where A_{22} is an $m \times m$ matrix, a_{33} is a scalar, and the remaining submatrices are consistent with that partitioning. (In some cases, A_{11} has zero rows and columns. You may verify that this does not invalidate the following calculations.) Choose the signature matrix

$$S = \begin{bmatrix} S_1 & 0 & 0 \\ 0 & S_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

such that $\left(a_{32} A_{22}^{-1} - a_{31}\right) S_1 \geq 0$ and $\left(-a_{32} A_{22}^{-1}\right) S_2 \geq 0$. Notice that, by the inductive hypothesis, $\det A_{22} > 0$ and therefore A_{22} is invertible. Now by assumption we have some

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} > 0$$

such that

$$SASx = \begin{bmatrix} S_1 A_{11} S_1 & S_1 A_{12} S_2 & S_1 a_{13} \\ S_2 A_{21} S_1 & S_2 A_{22} S_2 & S_2 a_{23} \\ a_{31} S_1 & a_{32} S_2 & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} > 0$$

Eliminating x_2, we obtain

$$\left(a_{33} - a_{32}A_{22}^{-1}a_{23}\right) x_3 = y_3 + \left(a_{32}A_{22}^{-1}a_{23} - a_{31}\right) S_1 x_1 - a_{32}A_{22}^{-1} S_2 y_2$$

Because of our choice of signature matrix, all terms on the right of this equation have nonnegative entries, and of course $x_3 > 0$. It follows that $a_{33} - a_{32}A_{22}^{-1}a_{23} > 0$. This means that

$$\det \begin{bmatrix} A_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} = \left(a_{33} - a_{32}A_{22}^{-1}a_{23}\right) \det A_{22} > 0$$

That is, this particular $(m+1) \times (m+1)$ principal minor is positive. The proof may be completed by permuting the rows and columns of A, so that the above argument applies to any $(m+1) \times (m+1)$ principal minor. It is easy to show that such permutations — provided that the column permutations are exactly the same as the row permutations — do not change the assumptions of the theorem, apart from changing the obvious array subscripts. □

This theorem gives conditions that are both necessary and sufficient for a matrix to have positive principal minors. In the following sections we will focus on conditions that are sufficient but easier to check.

3. Quasidominant matrices

Let us now look at a class of matrix that is very closely related to P-matrices.

Definition 22. *Let A be a square real matrix. Then A is called* quasidominant *iff there exists a vector $d > 0$ such that, for all i,*

$$d_i a_{ii} > \sum_{j \neq i} d_j \left|a_{ij}\right|$$

The relationship to diagonal dominance should be obvious. A matrix is quasidominant iff there exists a way to scale the columns that makes the result diagonally dominant. Less obviously, we can do the same thing by scaling the rows. The M-matrix test for quasidominance, to be given in the following section, implies that if A is quasidominant then so is A^T.

Quasidominant matrices have the following interesting property.

Theorem 61. *A real square matrix A is quasidominant iff there exists a vector $x > 0$ such that $SASx > 0$ for every signature matrix S.*

Proof. The condition $SASx > 0$ may be written as

$$a_{ii} x_i > -\sum_{j \neq i} s_{ii} s_{jj} a_{ij} x_j$$

where $s_{ii}s_{jj} = \pm 1$. If the inequality is to hold for every choice of the $s_{ii}s_{jj}$, then it is clearly equivalent to the condition

$$a_{ii} x_i > \sum_{j \neq i} \left|a_{ij}\right| x_j$$

which is the quasidominance condition. □

At first sight, it is difficult to see the difference between Theorem 60 and Theorem 61. The difference is this: in Theorem 61, the same x must work for any signature matrix S. The more general theorem permits a different x for each S. Comparing the two results, it is clear that all principal minors of a quasidominant matrix are positive; but the converse is not true. Note also that every symmetric quasidominant matrix must be positive definite.

We have not yet shown that these results have any relevance to systems theory. Our most important result is the following.

THEOREM 62. *If a square matrix F is quasidominant, then there exists a diagonal $P > 0$ such that $PF + F^T P > 0$.*

PROOF. Let $x > 0$ and $y > 0$ be such that $SFSx > 0$ and $SF^T Sy > 0$ for any signature matrix S. Let P be a diagonal matrix with diagonal entries $P_{ii} = y_i/x_i$. Then, using the fact that diagonal matrices commute with each other,

$$S\left(PF + F^T P\right) Sx = PSFSx + SF^T SPx = PSFSx + SF^T Sy > 0$$

This proves that $PF + F^T P$ is quasidominant, and therefore has positive principal minors. Since it is symmetric, it is positive definite. □

A similar result, but only for M-matrices — see below — was given by Tartar [**Tar71**] and Araki [**Ara75**].

4. M-matrices

Consider the set of real square matrices whose diagonal entries are positive and whose off-diagonal entries are nonpositive. M-matrices are a subset of this set: a matrix with this special sign pattern is an M-matrix iff all of its principal minors are positive. Equivalently, an M-matrix is a P-matrix with a special sign pattern. M-matrices have a number of interesting properties. Let us begin with one of the most basic results.

THEOREM 63. *If a real square matrix M has the sign pattern required of an M-matrix, then it is an M-matrix iff there exists a vector $x > 0$ such that $Mx > 0$.*

PROOF. If all principal minors of M are positive, the existence of a suitable x follows by setting $S = I$ in Theorem 60. For the converse, note that the condition $Mx > 0$ gives

$$x_i m_{ii} + \sum_{j \neq i} x_j m_{ij} > 0$$

But $m_{ij} \leq 0$ for all $j \neq i$, so that

$$x_i m_{ii} > \sum_{j \neq i} x_j \left|m_{ij}\right|$$

which is the definition of quasidominance. Since M is quasidominant, it follows from Theorem 61 that all of its principal minors are positive. □

This theorem has an obvious corollary.

THEOREM 64. *If a real square matrix M has the sign pattern required of an M-matrix, then it is an M-matrix iff it is quasidominant.*

PROOF. If M is an M-matrix, then there exists $x > 0$ such that $Mx > 0$. The reasoning in the proof of the previous theorem then shows that M is quasidominant. Conversely, if M is quasidominant, then all of its principal minors are positive. □

The relationship of M-matrices to quasidominant matrices is now obvious. For any real square matrix A, let a derived matrix $\widehat{A}$ be defined by $\widehat{A}_{ii} = A_{ii}$, and $\widehat{A}_{ij} = -\left|A_{ij}\right|$for $j \neq i$. Obviously A is quasidominant iff $\widehat{A}$ is quasidominant. It then follows from Theorem 64 that A is quasidominant iff $\widehat{A}$ is an M-matrix. This, in fact, is the easiest way to check a matrix for quasidominance.

It turns out that the inverse of an M-matrix is a matrix whose entries are all nonnegative. To show this, we first need a preliminary result.

LEMMA 17. *Let an M-matrix M be partitioned as*

$$M = \begin{bmatrix} M_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

where m_{22} is a scalar. Then $M_{11} - \frac{1}{m_{22}} m_{12} m_{21}$ is also an M-matrix.

PROOF. Since all principal minors of M are positive, then certainly m_{22} is positive. Let $x > 0$ be chosen such that $Mx > 0$. With the obvious partitioning, we have

$$\begin{aligned} M_{11}x_1 + m_{12}x_2 &= y_1 > 0 \\ m_{21}x_1 + m_{22}x_2 &= y_2 > 0 \end{aligned}$$

Eliminating x_2, we get

$$\left(M_{11} - \frac{1}{m_{22}} m_{12} m_{21} \right) x_1 = y_1 - \frac{1}{m_{22}} m_{12} y_2 > 0$$

where the final inequality comes from the fact that $-m_{12} \geq 0$. (Off-diagonal elements of an M-matrix are nonpositive.) The result then follows from Theorem 63. □

Now we can present the properties of an M-matrix the way they are usually presented.

THEOREM 65. *For a real square matrix M with $m_{ij} \leq 0$ for all $j \neq i$, the following conditions are equivalent.*

(1) All principal minors of M are positive.
(2) There exists a vector $x > 0$ such that $Mx > 0$.
(3) There exists a vector $y > 0$ such that $M^T y > 0$.
(4) M is nonsingular, and all entries of M^{-1} are nonnegative.

PROOF. The equivalence of 1 and 2 was shown in Theorem 63. To see that 4 implies 2, choose any $z > 0$ and let $x = M^{-1}z$; then obviously $x > 0$ if all entries of M^{-1} are nonnegative. By an almost identical argument, 4 implies 3. The proof that 1 implies 4 can be done by induction on the size of M. Suppose that we have shown that the result is true for M-matrices of size $k \times k$. (The proof for $k = 1$ is obvious.) Now consider a $(k+1) \times (k+1)$ matrix of the form

$$M = \begin{bmatrix} M_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

where m_{22} is a scalar. The inverse of M is given by

$$M^{-1} = \begin{bmatrix} A^{-1} & -\frac{1}{m_{22}} A^{-1} m_{12} \\ -\frac{1}{m_{22}} m_{21} A^{-1} & \frac{1}{m_{22}} (1 + \frac{1}{m_{22}} m_{21} A^{-1} m_{12}) \end{bmatrix}$$

where $A = M_{11} - \frac{1}{m_{22}} m_{12} m_{21}$. By Lemma 17, A is an M-matrix, and therefore invertible. By the inductive hypothesis, all entries of A^{-1} are nonnegative. Notice also, because of the sign pattern of an M-matrix, we have $m_{22} > 0$, $m_{12} \leq 0$, and $m_{21} \leq 0$. Clearly, then, all entries of M^{-1} are nonnegative. This concludes the proof. □

M-matrices were first introduced by Ostrowski. Alternative proofs of the equivalence of the four conditions may be found in [**Nik68**], or in the detailed coverage by Fiedler and Pták [**FP62**]. Araki [**Ara75**] was probably the first to show the relevance of M-matrices to systems and control theory.

In what follows, we will also need the following property.

LEMMA 18. *If $A - I$ is an M-matrix, then $I - A^{-1}$ is also an M-matrix.*

PROOF. Since the only difference between A and $A-I$ is that A has larger diagonal elements, A is also an M-matrix. That means that A is invertible, and all the elements of A^{-1} are nonnegative. Because $A-I$ is an M-matrix, then exists a vector $x>0$ such that $(A-I)x=u>0$. That means that $(I-A^{-1})x=A^{-1}u$. The right side is nonnegative because A^{-1} has nonnegative entries. Finally, $I-A^{-1}$ is nonsingular, because if it were singular then $A-I$ would be singular. □

One important application of these results to systems theory is Theorem 62. For a comparable finite gain result, we have the following.

LEMMA 19. *If A is a square matrix all of whose entries are nonnegative, and $I-A$ is an M-matrix, then there exists a diagonal $P>0$ such that $P-A^TPA$ is an M-matrix, and therefore $P-A^TPA>0$.*

PROOF. If $I-A$ is an M-matrix, then there exist vectors $x>0$ and $y>0$ such that $(I-A)x=u>0$ and $(I-A^T)y=v>0$. Define $P=\operatorname{diag}\{y_i/x_i\}$, so that $Px=y$. Then

$$\begin{aligned}(P-A^TPA)x &= y-A^TPAx=y-A^TPx+A^TP(I-A)x\\ &= v+A^Tu\end{aligned}$$

which is a positive vector because all entries of A are nonnegative. All off-diagonal elements of $P-A^TPA$ are nonpositive, therefore it is an M-matrix; and a symmetric M-matrix is positive definite. □

The usefulness of this result is a little restricted because of the restriction on the sign of the elements of A. Fortunately we can remove this restriction, as follows.

THEOREM 66. *If $I-|A|$ is an M-matrix, then there exists a diagonal $P>0$ such that $P-A^TPA>0$.*

PROOF. Note that

$$\left(P-A^TPA\right)_{ij}=\begin{cases}P_{ii}-\sum\limits_k A_{ki}P_{kk}A_{ki} & \text{if } j=i\\ -\sum\limits_k A_{ki}P_{kk}A_{kj} & \text{if } j\neq i\end{cases}$$

so that a sign change in any A_{ij} will have no effect on the diagonal elements, and will leave the off-diagonal elements no greater in magnitude than for the case where all the A_{ij} are positive. It follows that, if $P-|A|^T P|A|$ is an M-matrix, $P-A^TPA$ must be quasidominant.

If $I-|A|$ is an M-matrix, then from Lemma 19 there exists a diagonal P such that $P-|A|^T P|A|>0$. The sign pattern of this matrix shows that, if it is positive definite, it is an M-matrix. That means that the matrix $P-A^TPA$ is quasidominant. Since it is symmetric, it is positive definite. □

We can also get a result which is the symmetric partner of Lemma 19.

LEMMA 20. *If $A-I$ is an M-matrix, then there exists a diagonal $P>0$ such that $A^TPA-P>0$. In addition, $B^TPB-P>0$ for the same P and any B such that*

$$\begin{aligned}b_{ii}\geq a_{ii} &\qquad \textit{for all } i\\ 0\geq b_{ij}\geq a_{ij} &\qquad \textit{for all } j\neq i\end{aligned}$$

PROOF. If $A-I$ is an M-matrix, then from Lemma 18 $I-A^{-1}$ is an M-matrix. Lemma 19 then implies that there exists a diagonal $P>0$ such that $P-\left(A^{-1}\right)^T PA^{-1}>0$. Multiplying out the A^{-1}, we get $A^TPA-P>0$. The second part of the result follows from seeing that both A and B are M-matrices, that all elements of A^{-1} and B^{-1} are nonnegative, and that all elements of B^{-1}

are no greater than the corresponding elements of A^{-1}. The details may be found in [**Ara74**]. □

Our next two results are minor variants of Lemmas 19 and 20.

THEOREM 67. *If B is a real square matrix all of whose entries are nonnegative, and if $K > 0$ is a diagonal matrix sufficiently large that $K - B$ is an M-matrix, then there exists a diagonal $D > 0$ such that $B^T DB < DK^2$.*

PROOF. From the definition of quasidominance, or otherwise, it is easy to see that $I - K^{-1}B$ is an M-matrix. From Lemma 19, there exists a diagonal $P > 0$ such that $P - B^T K^{-1} P K^{-1} B > 0$. Setting $D = PK^{-2}$, we have the desired result. □

THEOREM 68. *Let $(-B)$ be an M-matrix, and let $K > 0$ be any diagonal matrix such that $(-B - K)$ is also an M-matrix. Then there exists a diagonal $D > 0$ such that $B^T DB - DK^2 > 0$. Further, for the same D and K we have*

$$\tilde{B}^T D\tilde{B} - DK^2 > 0$$

for any $\tilde{B} = \left[\tilde{b}_{ij}\right]$ that satisfies

$$\begin{aligned} \tilde{b}_{ii} \leq b_{ii} < 0 \qquad & \textit{for all } i \\ 0 < \tilde{b}_{ij} \leq b_{ij} \qquad & \textit{for all } j \neq i \end{aligned}$$

PROOF. Reasoning as before, $(-K^{-1}B - I)$ is an M-matrix, so from Lemma 20 there exists a diagonal $P > 0$ such that $\left(K^{-1}B\right)^T PK^{-1}B - P > 0$. Setting $D = PK^{-2}$, we have the first part of the result. For the remainder, observe that the matrix $\left(-K^{-1}\tilde{B}\right)$ satisfies the conditions of the corresponding part of Lemma 20. □

5. Transformations

From the preceding results, it is clear that we often want to find a diagonal $P > 0$ such that a certain condition is satisfied. (It would be easier if P did not have to be diagonal, but we do not have that luxury.) The conditions provided by our theorems are sufficient but not necessary conditions. Can we do better?

It is easy to show that if we define $F = (I - A)(I + A)^{-1}$, or conversely $A = (I + F)^{-1}(I - F)$, then $PF + F^T P > 0$ if and only if $P - A^T PA > 0$. That means that if a matrix fails to satisfy one of our sufficiency tests, we have the option of applying a transformation and then checking a different test. If that too fails, we can try other transformations like $F = (I + A)(I - A)^{-1}$ or $A = e^F$.

More complex inequalities can also be reduced to one of the basic forms $PF + F^T P > 0$ or $P - A^T PA > 0$ by simple transformations. For example, the inequality

$$X^T P\Sigma X - Y^T P\Sigma Y > 0$$

(where P and Σ are both diagonal) can be reduced to a simpler form by setting $F = \Sigma(X - Y)(X + Y)^{-1}$. If Σ is a unit matrix and X is invertible, a different simplification appears after setting $A = YX^{-1}$.

Bibliography

[ABJ75] B.D.O. Anderson, N.K. Bose, and E.I. Jury. Output feedback stabilization and related problems—solution via decision methods. *IEEE Trans Automat Contr*, AC-20:53–66, Feb 1975.

[AK72] M. Araki and B. Kondo. Stability and transient behaviour of composite nonlinear systems. *IEEE Trans Automat Contr*, AC-17:537–541, 1972.

[AM75] B.D.O. Anderson and P.J. Moylan. Structure result for nonlinear passive systems. In *Int Symp Operator Theory of Networks and Syst*, Montreal, Canada, Aug 1975.

[And75] B.D.O. Anderson. Corrections to: Algebraic properties of minimal degree spectral factors. *Automatica*, 11:321–322, 1975.

[Ara74] M. Araki. M-matrices (Matrices with nonpositive off-diagonal elements and positive principal minors). Technical Report 74/19, Dept Computing and Control, Imperial College of Science and Technology, London, Mar 1974.

[Ara75] M. Araki. Application of M-matrices to the stability problems of composite dynamical systems. *J Math Anal Appl*, 52:309–321, 1975.

[Ara76] M. Araki. Input-output stability of composite feedback systems. *IEEE Trans Automat Contr*, AC-21:254–259, Apr 1976.

[AV73] B.D.O Anderson and S. Vongpanitlerd. *Network Analysis and Synthesis — A Modern Systems Theory Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[BS67] A.R. Bergen and M.A. Sapiro. The parabola test for absolute stability. *IEEE Trans Autom Contr*, AC-12:312–314, 1967.

[CN68] Y.S. Cho and K.S. Narendra. An off-axis circle criterion for the stability of feedback systems with a monotonic nonlinearity. *IEEE Trans Autom Contr*, AC-13:413–416, 1968.

[Coo74] P.A. Cook. On the stability of interconnected systems. *Int J Contr*, 20(3):422–427, 1974.

[FP62] M. Fiedler and V. Pták. On matrices with non-positive off-diagonal elements and positive principal minors. *Czech Math J*, 12:382–400, 1962.

[GN65] David Gale and Hukukane Nikaidô. The Jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159:81–93, 1965.

[HM75] D.J Hill and P.J. Moylan. Cyclo-dissipativeness, dissipativeness and losslessness for nonlinear dynamic systems. Technical Report Tech Rept EE7526, Univ Newcastle, New South Wales, Nov 1975.

[HM76] D.J. Hill and P.J. Moylan. The stability of nonlinear dissipative systems. *IEEE Trans Automat Contr*, AC-21(5):708–711, Oct 1976.

[HM77] D.J. Hill and P.J. Moylan. Stability results for nonlinear feedback systems. *Automatica*, 13:377–382, 1977.

[HM78] D.J. Hill and P.J. Moylan. A general result on the instability of feedback systems. In *IEEE Int Symp on Circuits and Syst*, New York, May 1978.

[HM80a] D.J. Hill and P.J. Moylan. Connections between finite gain and asymptotic stability. *IEEE Trans Automat Contr*, AC-25(5):931–936, Oct 1980.

[HM80b] D.J. Hill and P.J. Moylan. Definitions of passivity for nonlinear circuits and systems. In *Proc 1980 IEEE Symp on Circuits and Syst*, pages 856–859, Apr 1980.

[HM80c] D.J. Hill and P.J. Moylan. Dissipative dynamical systems: basic input-output and state properties. *J Franklin Inst*, 309(5):327–357, May 1980.

[HM83] D.J. Hill and P.J. Moylan. General instability results for interconnected systems. *SIAM J Control and Optimization*, 21(2):256–279, March 1983.

[Jac74] N. Jacobson. *Basic Algebra I*. San Francisco: W.H. Freeman, 1974.

[MA73] P.J. Moylan and B.D.O. Anderson. Nonlinear regulator theory and an inverse optimal control problem. *IEEE Trans Automat Contr*, AC-18(5):460–464, Oct 1973.

[MCS82] P.J. Moylan, L.O. Chua, and E.W. Szeto. When is a device passive? *Internat J Circuit Theory and Appln*, 10:151–165, Feb 1982.

[MH78] P.J. Moylan and D.J. Hill. Stability criteria for large-scale systems. *IEEE Trans Automat Contr*, AC-23(2):143–149, Apr 1978.

[MH79] P.J. Moylan and D.J. Hill. Tests for stability and instability of interconnected systems. *IEEE Trans Automat Contr*, AC-24(4):574–575, Aug 1979.

[Mic74] A.N. Michel. Stability analysis of interconnected systems. *SIAM J Contr*, 12(3):554–579, Aug 1974.

[Moy74] P.J. Moylan. Implications of passivity in a class of nonlinear systems. *IEEE Trans Automat Contr*, AC-19(4):373–381, Aug 1974.

[Moy75] P.J. Moylan. On a frequency-domain condition in linear optimal control theory. *IEEE Trans Automat Contr*, AC-20:806, Dec 1975.

[Moy77] P.J. Moylan. Matrices with positive principal minors. *Linear Algebra and Its Applications*, 17:53–58, 1977.

[Moy80] P.J. Moylan. A connective stability result for interconnected passive systems. *IEEE Trans Automat Contr*, AC-25(4):812–813, Aug 1980.

[Moy81] P.J. Moylan. A relaxed dissipativeness test for the stability of large-scale systems. *IEEE Trans Automat Contr*, AC-26(2):545–547, Apr 1981.

[Moy82] P.J. Moylan. Circle criteria revisited. In *Proc IE Aust 2nd Conf on Control Engg*, pages 56–58, Newcastle, Aug 1982.

[Moy86] P.J. Moylan. Dissipativeness concept in stability. In *Encylopedia of Systems and Control*. Pergamon Press, Jun 1986.

[MP72] A.N. Michel and D.W. Porter. Stability analysis of composite systems. *IEEE Trans Automat Contr*, AC-17:222–226, Apr 1972.

[MVV80] P.J. Moylan, A. Vanelli, and M. Vidyasagar. On the stability and well-posedness of interconnected nonlinear dynamical systems. *IEEE Trans Circuits Syst*, CAS-27(11):1097–1101, Nov 1980.

[Nik68] H. Nikaidô. *Convex Structures and Economic Theory*. New York: Academic, 1968.

[PM74] D.S. Porter and A.N. Michel. Input-output stability of time-varying nonlinear multiloop feedback systems. *IEEE Trans Automat Contr*, AC-19:422–427, Aug 1974.

[PM93] H.R. Pota and P.J. Moylan. Stability of locally dissipative interconnected systems. *IEEE Trans Automat Contr*, AC-38(2):308–312, 1993.

[Pop61] V.M. Popov. Absolute stability of nonlinear systems of automatic control. *Avt. i Telemekh.*, 22(8):961–980, 1961. (in Russian).

[Pop73] V.M. Popov. *Hyperstability of Automatic Control Systems*. Berlin: Springer, 1973.

[RM76] R.D. Rasmussen and A.N. Michel. Stability of interconnected dynamical systems described on Banach spaces. *IEEE Trans Automat Contr*, AC-21:464–471, Aug 1976.

[Ros70] H.H. Rosenbrock. *State Space and Multivariable Theory*. London, England: Nelson, 1970.

[San64] I.W. Sandberg. On the L_2 boundedness of solutions of nonlinear functional equations. *Bell System Tech J*, 43:1581–1599, 1964.

[Sti15] E. Stiemke. Über positive Lösungen homogener linearer Gleichungen. *Mathematische Annalen*, 76:340–342, 1915.

[Tar71] L. Tartar. Une nouvelle caracterisation des M matrices. *Ref Fr Inf Rech Opér*, 5:127–128, 1971.

[Vid79a] M. Vidyasagar. L_2-stability of interconnected systems using a reformulation of the passivity theorem. *IEEE Trans Circuits Syst*, CAS-24:637–645, 1979.

[Vid79b] M. Vidyasagar. New passivity-type criteria for large-scale interconnected systems. *IEEE Trans Automat Contr*, AC-24(4):575–579, August 1979.

[Š72] D. Šiljak. Stability of large-scale systems. In *Proc IFAC 5th World Congr, Part IV*, pages C–32:1–11, Paris, France, Jun 1972.

[Š76] D.D. Šiljak. Large-scale systems: complexity, stability, reliability. *J Franklin Inst*, 301:49–69, Jan-Feb 1976.

[Š78] D. Šiljak. *Large-Scale Dynamic Systems*. North-Holland, The Netherlands, 1978.

[Wil71] J.C. Willems. Least squares stationary optimal control and the algebraic riccati equation. *IEEE Trans Automat Contr*, AC-16:621–634, Dec 1971.

[Wil72] J.C. Willems. Dissipative dynamical systems, part I: General theory; part II: Linear systems with quadratic supply rates. *Arch. Rational Mech. Anal.*, 45:321–393, 1972.

[Wil74] J.C Willems. On the existence of a nonpositive solution to the riccati equation. *IEEE Trans Automat Contr*, AC-19:592–593, Oct 1974.

[Zam66] G. Zames. On the input-output stability of time-varying nonlinear feedback systems, part I: conditions derived using concepts of loop gain, conicity, and positivity; part II: conditions involving circles in the frequency plane and sector nonlinearities. *IEEE Trans Automat Contr*, AC-11:228–238,465–476, 1966.

Index