DIFFERENTIAL INEQUALITIES AND THE

MATRIX RICCATI EQUATION

by

A.N. Stokes

A thesis presented to the
Australian National University
for the degree of
Doctor of Philosophy
Canberra
November, 1972

STATEMENT

This thesis is my own work. Results which are not my own are indicated in the text.

A. N. Stokes

ACKNOWLEDGEMENTS

My work for this thesis was supervised by Mr W.A. Coppel. I am grateful to him for his many helpful suggestions, his constructive criticism and his generous assistance during the preparation of this thesis.

I also thank the other people at the Australian National University with whom I worked. Some of the material in this thesis developed from two series of seminars, led by Mr Coppel; Dr A. Howe, Dr G.C. O'Brien and Mr J. Davy were also regular participants.

I thank Mrs Barbara Geary, both for her typing, and for her work in deciphering the manuscript drafts of this thesis.

I am grateful to the Australian National University for awarding me a research scholarship to carry out this work.

CONTENTS

INTRODUCTION

The matrix Riccati equation has attracted attention recently because of its occurrence in a number of different situations. Its solutions determine solutions of the optimal linear regulator problem (Kalman [*1*], Athans and Falb [*1*]); the existence of solutions on an interval is related to disconjugacy of a linear Hamiltonian system on an interval (Reid [*14*], Coppel [*3*]) and Schumitzky [*1*] has demonstrated an equivalence between solutions of matrix Riccati equations and Fredholm resolvents. Most recently, Fair [*1*] has written about continued fraction solutions of a Riccati equation in a Banach algebra.

In this thesis, only symmetric matrix Riccati equations, of the form

$$R[W] = W' + A(t) + B^*(t)W + WB(t) + WC(t)W = 0 \ , \qquad (1)$$

are considered where $A(t)$, $B(t)$, $C(t)$ are continuous $n \times n$ matrix functions of t, and $A(t)$ and $C(t)$ are symmetric. Only symmetric matrix solutions $W(t)$ are considered.

One property of (1), of which full use has not always been made, is that it preserves the ordering of solutions. For symmetric matrices A and B, we say $A \geq B$ if $A - B$ is nonnegative definite. A more general and exact statement about this order-preserving property is that if $W_1(t)$, $W_2(t)$ are symmetric matrix functions differentiable on an interval $[a, b]$, and $W_1(a) \geq W_2(a)$, $R\left[W_1(t)\right] \geq 0$ and $R\left[W_2(t)\right] \leq 0$, then $W_1(b) \geq W_2(b)$. Or, if $W_1(b) \geq W_2(b)$, $R\left[W_1(t)\right] \leq 0$, $R\left[W_2(t)\right] \geq 0$, then $W_1(a) \geq W_2(a)$.

These statements are like those proved in the theory of differential inequalities for certain vector systems, with a component-wise vector ordering, and indeed Coppel [*3*] has shown that similar methods of proof can be used for symmetric matrix Riccati systems as for vector systems (Coppel [*2*]).

In Chapter 1, we generalise the usual arguments so that they apply in a vector space where the ordering is abstractly defined, and the resulting differential inequalities are seen as consequences of the axioms of this definition. Some examples of the use of different orderings are given; in particular the positive cone can be the set of vectors with positive coefficients (component-wise ordering), or a circular cone (a Lorentz-type ordering), or the cone corresponding to the set of positive-definite matrices. The latter cone is investigated in the second chapter, where a theorem is derived about inequalities, which is extensively applied in the last three chapters.

A consequence of this geometric type of approach is that the Riccati equation has a special status in multi-dimensional systems, being the only type of symmetric matrix equation which preserves the ordering of solutions as t increases or decreases.

If the set of solutions of (1) existing on an interval can be ordered, it makes sense to speak of a maximal and minimal element of such a set. If $C(t) \geq 0$ we show that there are two difficulties in the way of proving the existence of such elements on an interval I :

a) there may be no solutions at all existing on I ;

b) the maximum, or minimum, solution may be infinite-valued.

Solutions of the Riccati equation (1) correspond to solutions of the Hamiltonian system

$$Y' = B(t)Y + C(t)Z \ ,$$

$$Z' = -A(t)Y - B^*(t)Z \ , \quad Y, Z \quad n \times n \text{ matrices,} \qquad (2)$$

in the following way: if $Y(t)$ is a solution of

$$Y' = \big(B(t)+C(t)W(t)\big)Y$$

where $W(t)$ is a solution of (1), then $\langle Y(t), W(t)Y(t)\rangle$ is a solution of (2), and if $\langle Y(t), Z(t)\rangle$ is a solution of (2) with $Y(t)$ invertible on some interval, then $Z(t)Y^{-1}(t)$ is a solution of (1).

In this sense, solutions of (2) corresponding to maximal solutions (or minimal solutions) of (1) on an interval are called principal solutions of (2). Reid [14], Hartman [7] and Coppel [3] use a different definition, and in these, and other developments of principal solutions, conditions like controllability are imposed on the coefficients to ensure that b) does not arise. Difficulty a) is circumvented by the assumption of disconjugacy.

In Chapter 3, we approach the Riccati equation directly, using differential inequalities, to show that a maximal element of the set of solutions existing on I can be found, provided the set is non-empty. The maximal element may be infinite-valued, but most of the implications for the structure of the set of solutions remain unchanged. In particular a principal solution for (2) can still be found, and has most of the properties of principal solutions under the more restrictive conditions of previous writers, but may be singular if the maximal solution is infinite valued.

A linear Hamiltonian system

$$y' = B(t)y + C(t)z \ ,$$

$$z' = -A(t)y - B^*(t)z \ , \quad y, z \quad n \times 1 \quad \text{vectors}, \tag{3}$$

with A, B, C coefficients as for (1) and (2), is disconjugate on an interval I if, whenever $[a, b] \subseteq I$, and $\langle y, z \rangle$ is a solution of (3), then $y(a) = y(b) = 0$ only if $y(t) = 0$ on $[a, b]$.

In Chapter 3, we extend this definition slightly so that neither of the two principal solutions generate solutions with zeros; then we show that this extended disconjugacy definition is necessary and sufficient for the existence of a solution of (1), and so for the properties that follow.

These are the two main results of Chapter 3. A necessary condition for disconjugacy is also proved which incorporates some previous results and gives an upper limit for the length of interval with a certain point as end-point on which (2) can be disconjugate. Also an example involving Lorentz ordering of a vector space, introduced in Chapter 1, is pursued to show closely analogous arguments concerning maximal solutions.

In Chapter 4 a continued fraction expansion associated with solutions of a Riccati equation is given. It is shown that the convergents form good approximations near the point about which the expansion is made, but not that the fraction converges. However it is shown that the sequence of convergents is an improving sequence of bounds to a solution. Based on this, a sequence of increasingly critical necessary conditions for disconjugacy (or oscillation criteria), is given, and also a criterion requiring an assumption of positivity about the sign of only one coefficient.

Chapter 5 is concerned with asymptotic behaviour of the Riccati equation and the associated linear system. We show that the arguments used to prove exponential stability for certain solutions of the uniformly observable and controllable linear regulator problem can be used to show stability of some specifiable sort in many other cases. Deductions are made about the tendency of solutions of the Riccati equation to aggregate at infinity.

CHAPTER 1

GENERAL DIFFERENTIAL INEQUALITIES

Introduction

The theory of differential inequalities is concerned with the question: if two solutions $y_1(t)$, $y_2(t)$ of the nonlinear system

$$y' = f(t, y) \tag{1}$$

are in some inequality relationship at a certain point (that is, $y_1(b) \geq y_2(b)$ or $y_1(b) < y_2(b)$ etc), is this relationship preserved as t increases or decreases? In the case where y, f are scalars, the answer is generally yes, subject to special definitions where solutions are not uniquely determined by their initial values. Otherwise, f must fulfil a special condition. In this case the theory also allows comparison of the solutions of (1) with solutions of inequalities like $\frac{dy}{dt} \geq f(t, y)$.

Differential inequalities have normally referred only to inequalities defined by the usual partial ordering of a vector space, where one vector exceeds another if all its respective components are greater. This chapter shows that the arguments used in that case apply when the partial ordering is abstractly defined, using any non-degenerate positive cone of vectors.

There is one exception; the proof of a theorem involving assertion of existence of solutions (Theorems 5 and 6) does not carry over if solutions are not uniquely determined by their initial values.

The exposition below broadly parallels that of Coppel [*2*, Chapter 1]. Our Theorems 1 to 3 correspond to basic theorems in that reference. Other expositions of the usual theory are contained in Szarski [*1*], Walter [*1*], Lakshmikantham and Leela [*1*].

As examples of applications, there is a proof of a uniqueness result, and a list of some cases where inequalities involving non-standard orderings are useful.

Preliminaries and the type K condition

DEFINITIONS. A cone K in R^n is a convex set with the properties that if $a \in K$ then $ta \in K$ iff $t \geq 0$, unless $a = 0$. Also $0 \in K$.

Let C be a closed cone in R^n with a non-empty interior. The reason for this latter proviso is indicated in the remark following Theorem 1. An ordering in R^n is defined thus: $x \geq y$ if $x-y \in C$, and $x > y$ if $x-y \in C^i$, where C^i is the interior of C . C is referred to as the positive cone.

A dual cone C^* is defined: $x \in C^*$ iff $(x, y) \geq 0$ for all y in C , where (x, y) is the scalar product of x and y .

LEMMA 1. $C^{**} = C$.

Proof. If $x \in C$, then $(x, y) \geq 0$ for all y in C^* . Therefore $x \in C^{**}$ so $C^{**} \supseteq C$.

The converse inclusion is a consequence of the duality theorem for convex sets, which says that a closed convex set is equal to the intersection of all half-spaces containing it. For a proof, see Luenberger [1, p. 215, Proposition 1].

The special condition that is needed for all the later results is given by

DEFINITION. A function $f(x)$ with both its domain S and its range in R^n , is of type $K_+\{K_-\}$ in S if, whenever for two points x and y in S , $x \geq y$ and $(x, z) = (y, z)$ for some z in C^* , then $\big(f(x), z\big) \geq \big(f(y), z\big)\{\big(f(x), z\big) \leq \big(f(y), z\big)\}$.

For brevity, the function $f(t, x)$ and the equation $x' = f(t, x)$ will also be said to be of type K_+ or K_- on some domain D in $R \times R^n$ if $f(t, x)$ is respectively of type K_+ or K_- for each t in the domain of points (t, x) in D . Also for brevity, if for each t in an interval J , there is a non-empty set of vectors $x : (t, x) \in D$, then we say D contains J .

The most general form of the differential inequalities to come involves the use of one-sided upper and lower derivatives. In the general case these derivatives will not correspond to specific vectors, because pairs or sets of vectors may not have least upper bounds.

However the overall inequalities can be made meaningful, in the following sense: for a function $y(t)$ in R^n, $D^+y(t) \overset{\geq}{\{>\}} c$, $c \in R^n$, means that $D^+(y(t), z) \overset{\geq}{\{>\}} (c, z)$ for all vectors z in C^*, $z \neq 0$, that is

$$\overline{\lim_{\substack{u>t\\u\to t}}}\, (y(u)-y(t), z) \overset{\geq}{\{>\}} (c, z) \quad \forall z \in C^*\,, \quad z \neq 0\,.$$

With this usage both D^+ and $\geq$ have artificial meanings.

The expressions $D^-y \geq c$, $D_+y \geq c$, $D_-y \geq c$ are correspondingly defined in terms of their scalar equivalents.

Basic theorems about inequalities

THEOREM 1. *Let $f(t, x)$ be continuous and of type K_+ in some domain D containing an interval $[a, b]$. Let $y(t)$, $z(t)$ be continuous functions on $[a, b]$ whose graphs are in D and which satisfy on $(a, b]$ the inequalities*

$$D^-y > f(t, y)\,, \quad D_-z \leq f(t, z)\,, \quad \text{and} \quad y(a) > z(a)\,.$$

Then $y(t) > z(t)$ on $[a, b]$.

The theorem applies particularly when $z(t)$ is a solution of

$$z' = f(t, z)\,. \tag{1}$$

Proof. By continuity $y(t) > z(t)$ on some interval $[a, a+d]$ where $d > 0$.

If the inequality $y(t) > z(t)$ does not hold throughout $[a, b]$ there would exist a point c, $a \leq c \leq b$, for which $y(t) > z(t)$ on $[a, c)$, $y(c) \geq z(c)$, and $(y(c)-z(c), x) = 0$ for at least one vector x in C^*, $x \neq 0$.

Then

$$\begin{aligned} D^-(y(c),\, x) &> (f(c,\, y(c)),\, x) \\ &\ge (f(c,\, z(c)),\, x) \quad \text{since } f \text{ is of type } K_+ \\ &\ge D_-(z(c),\, x)\ . \end{aligned}$$

Since $(y(c), x) = (z(c), x)$ it follows that for certain values of t less than and arbitrarily close to c, we have $(y(t), x) < (z(t), x)$ contradicting the definition of c. So $y(t) > z(t)$ on $[a, b]$.

There are three analogous theorems:

A. *If it is the lower function which satisfies a strict inequality, that is* $D^- y \ge f(t, y)$, $D_- z < f(t, z)$ *on* $(a, b]$, *the conclusion that* $y(t) > z(t)$ *remains true.*

B. *If* $f(t)$ *is of type* K_-, *and* $y(t)$, $z(t)$ *are continuous functions satisfying* $D^+ y > f(t, y)$, $D_+ z \le f(t, z)$ *on* $[a, b)$, *and* $y(b) < z(b)$, *then* $y(t) < z(t)$ *on* $[a, b]$.

C. *Again, in* B *the strict inequality may apply to* z *rather than* y.

Remark. If C has an empty interior, then there are no vectors in a strict inequality relationship to each other, so Theorem 1 is vacuously true. But then it is useless for the purpose it serves in Theorem 3, where weak inequality relations are derived as the limits of strong inequality relations.

DEFINITION. We define a solution of the differential system

$$x' = f(t, x) \tag{1}$$

on an interval I to be a right maximal solution if for every $t_0 \in I$ any solution $x(t)$ of (1) such that $x(t_0) \le \tilde{x}(t_0)$ satisfies the inequality $x(t) \le \tilde{x}(t)$ for all $t > t_0$ in I for which $x(t)$ is defined.

THEOREM 2. *Let* $f(t, x)$ *be continuous and of type* K_+ *in an open set* D. *Then the differential equation* (1) *has a unique right maximal solution passing through any point* (t_0, ξ_0) *of* D, *which is defined in an interval* $[t_0, \bar{t})$ *and tends to the boundary of* D *as* $t \to \bar{t}$.

Proof. If a right maximal solution exists, it must clearly be unique. Choose a vector ε in C^l and let $\varphi_n(t)$ be any solution of the initial value problem:

$$x' = f(t, x) + \varepsilon/n \ ,$$
$$x\left(t_0\right) = \xi_0 + \varepsilon/n \ ,$$

n a positive integer.

There is an interval $\left[t_0, t_1\right]$ of positive length throughout which the functions $\varphi_n(t)$ are defined and have their graphs in a prescribed neighbourhood of $\left(t_0, \xi_0\right)$ for all sufficiently large n. By Theorem 1, $\varphi_n(t) < \varphi_m(t)$ if $n > m$. Since the sequence $\{\varphi_n\}$ is equicontinuous in t and decreasing in n it converges uniformly on the interval $\left[t_0, t_1\right]$ as $n \to \infty$, and the limit function $\psi(t)$ is a solution of (1) passing through $\left(t_0, \xi_0\right)$.

If $x(t)$ is a solution of (1) such that $x\left(t_2\right) \le \psi\left(t_2\right)$ where $t_0 \le t_2 \le t_1$ then $x\left(t_2\right) < \varphi_n\left(t_2\right)$ for all large n and hence, by Theorem 1, $x(t) < \varphi_n(t)$ for $t_2 < t \le t_1$. Letting $n \to \infty$, we get $x(t) \le \psi(t)$ for $t_2 < t \le t_1$.

By the uniqueness of the right maximal solution there exists a right maximal solution $\tilde{\psi}(t)$ through $\left(t_0, \xi\right)$ which is not continuable as a right maximal solution. Let $\bar{t}$ be the right end-point of its interval of definition. If the graph of $\tilde{\psi}(t)$ had a limit point $(\bar{t}, \bar{\xi})$ inside D, we would have $\tilde{\psi}(t) \to \tilde{\xi}$ for $t \to \bar{t}$ [see for example, Hartman [2], Chapter II, Theorem 3.1]. But then, by what we have proved, $\bar{\psi}(t)$ could be continued past $\bar{t}$ as a right maximal solution.

Analogues of Theorem 2: *An equivalent argument establishes the existence of a right minimal solution. If* f *is of type* K_-, *then there exist left maximal and left minimal solutions.*

THEOREM 3. *Let* $f(t, x)$ *be continuous and of type* K_+ *in an open set* D. *Let* $x(t)$ *be a right maximal solution of* (1) *on an interval* $[a, b]$. *If* $z(t)$ *is continuous on* $[a, b]$, *satisfies* *the differential inequality* $D_{-}z \leq f(t, z)$ *on* $(a, b]$ *and* $z(a) \leq x(a)$ *, then* $z(t) \leq x(t)$ *for* $a \leq t \leq b$ *.*

Proof. Let c be the greatest value of t such that $z(s) \leq x(s)$ for $a \leq s \leq t$ and suppose, contrary to the theorem, that $c < b$. Choose a vector $\varepsilon > 0$ and let $\varphi_n(t)$ be a solution of the initial value problem

$$x' = f(t, x) + \varepsilon/n$$

$$x = x(c) + \varepsilon/n \quad \text{for} \quad t = c$$

in an interval $[c, c+\delta]$. By the proof of Theorem 2, $\varphi_n(t)$ converges to $x(t)$ on this interval as $n \to \infty$. On the other hand, by Theorem 1, $z(t) < \varphi_n(t)$ for $c \leq t \leq c+\delta$. Letting $n \to \infty$ we get $z(t) \leq x(t)$ for $c \leq t \leq c+\delta$. This contradicts the definition of c .

Analogous forms: A. If $x(t)$ is a right minimal solution and $y(t)$ satisfies $D^{-}y \geq f(t, y)$ on $(a, b]$ and $y(a) \geq x(a)$, then $y(t) \geq x(t)$ on $[a, b]$.

B. If f is of type K_{-} , $x(t)$ is a left maximal solution, and $z(t)$ satisfies $D_{+}z \geq f(t, z)$ on $[a, b)$ and $z(b) \leq x(b)$, then $z(t) \leq x(t)$ on $[a, b]$.

C. If f is of type K_{-} , $x(t)$ is a left minimal solution and $y(t)$ satisfies $D^{+}y \leq f(t, y)$ on $[a, b)$ and $y(b) \leq x(b)$ then $y(t) \leq x(t)$ on $[a, b]$.

COROLLARY. *A continuous vector function* $z(t)$ *is non-increasing on an interval* $[a, b]$ *iff it satisfies the differential inequality* $D_{-}z \leq 0$ *(or* $D_{+}z \leq 0$ *) on* (a, b) *.*

Proof. The necessity of the condition follows from the definition of D_{-} . For if $z(t_2) - z(t_1) \in C$ whenever $b \geq t_2 \geq t_1 \geq a$, then for any $x \in C^*$, $(z(t_2), x) \geq (z(t_1), x)$ so $D_{-}(z(t), x) \leq 0$ and $D_{+}(z(t), x) \leq 0$ and by definition $D_{-}z \leq 0$ on $[a, b]$.

The sufficiency of the condition follows from Theorem 3 with $f \equiv 0$.

Remarks. The introduction of maximal solutions in Theorems 2 and 3 is necessary to make meaningful statements about the relationship of solutions which are not uniquely determined at a given point. If solutions of (1) are uniquely determined by their initial values, then Theorems 1 and 3 assert that if f is of type K_+, all inequality relations among solutions of (1) are preserved as t increases, whereas if f is of type K_-, all relations are preserved as t decreases. If f is both of type K_+ and K_-, (again assuming uniqueness) then one can meaningfully talk of solutions being ordered over a whole interval, and this ordering is just the ordering of their values at a single point arbitrarily chosen.

This is quite a strong statement, and the following theorem shows that the respective type K conditions are both necessary and sufficient for some of the results deduced from them.

All scalar functions are of type K_+ and K_-, but for functions of n-vectors, the conditions are quite restrictive. Szarski [*1*], for example, shows that a system (1) both of type K_+ and K_- in the usual vector ordering reduces to a degenerate system of n separate equations each in one variable. In Chapter 2, we show that the corresponding restriction to a system with the positive definiteness ordering of symmetric matrices reduces to the much more interesting matrix Riccati equation.

Here is the demonstration of the necessity of type K conditions:

THEOREM 4. *Let $f(t, x)$ be defined in an open set D of $R \times R^n$. Suppose that, whenever there are two points (t, x_0), (t, y_0) in D for which $x_0 \geq y_0$, then on some non-empty interval $[t, u]$ there exist in D solutions of*

$$D^+y \geq f(s, y), y(t) = y_0$$

$$D_+x \leq f(s, x), x(t) = x_0$$

for which $x(s) \geq y(s)$ on $[t, u]$. Then $f(t, x)$ is of type K_+ on D.

Proof. Let x_0, y_0 be any two points as above for which there is a vector z in C^* : $(x_0, z) = (y_0, z)$. Then for all s in $[t, u]$, $(x(s)-y(s), z) \geq 0$, so $D_+(x(t)-y(t), z) \geq 0$. But

$$D_+(x(t), z) \leq (f(t, x(t)), z)$$

$$D^+(y(t), z) \geq (f(t, y(t)), z)$$

so

$$(f(t, x_0), z) \geq (f(t, y_0), z) .$$

Since t, z and x_0, y_0 are arbitrary, subject to $x_0 \geq y_0$, then f is of type K_+ in D .

COROLLARY 1. *If t is replaced by $-t$ in the statement, with consequent modifications of the inequalities, then $f(t, x)$ is of type K_- in D .*

COROLLARY 2. *With the stronger hypothesis that whenever (t, x_0) and $(t, y_0) \in D$, and $x_0 \geq y_0$ then on some interval (a, b) containing t there exist solutions $x(u)$ and $y(u)$ of* (1) *such that $x(t) = x_0$, $y(t) = y_0$ and $x(u) \geq y(u)$ in (a, b) , the result is that $f(t, x)$ is both of type K_+ and K_- in D .*

Existence of solutions

With the usual vector ordering of R^n , the existence of two solutions $y(t), z(t)$ of the inequalities $D^-y \geq f(t, y)$, $D_-z \geq f(t, z)$ on $(a, b]$, with $y(a) \geq z(a)$, ensures that whenever $y(a) \geq x \geq z(a)$, there is a solution of (1) with $x(a) = x$ existing and constrained to lie between $y(t)$ and $z(t)$ on $[a, b]$. [Coppel [1], p. 30.]

In the case where solutions of (1) are uniquely determined, the proof of the corresponding proposition in our situation is a consequence of Theorem 2 - a solution is continuable until it approaches the boundary of its domain of definition. This is set out in Theorem 6 below. But if solutions are not unique, then the more subtle proof given by Coppel for the proposition set out above does not carry over directly. However an analogous approach can be made to work provided the inequalities are strict inequalities, as in Theorem 5 below.

THEOREM 5. *Let* $f(t, x)$ *be a continuous function defined on a domain* D *in* $R \times R^n$ *which includes an interval* $[a, b]$ *. Let* $y(t)$, $z(t)$ *be continuous functions for which* $D^-y > f(t, y)$ *,* $D_-z < f(t, z)$ *,* $a < t \le b$ *,* $z(a) < y(a)$ *and if* $K(t) = \{u : y(t) \ge u \ge z(t)\}$ *, then* $(t, K(t)) \subset D$ *.*

Then for any $x_0 : y(a) > x_0 > z(a)$ *, the initial value problem*

$$\begin{aligned} x' &= f(t, x) \\ x(a) &= x_0 \end{aligned} \tag{1}$$

has a solution $x(t)$ *which is defined and satisfies the inequalities* $z(t) < x(t) < y(t)$ *on* $[a, b]$ *.*

Proof. For each t , $K(t)$, being the intersection of two closed cones, is closed. Let $S = \{(t, u) : a \le t \le b, u \in K(t)\}$. Then S is closed, since $K(t)$ is continuous.

By Theorem 1, $y(t) > z(t)$ on $[a, b]$, and any solution $x(t)$ of (1) with $x(a) = x_0$ lies in the interior of S wherever it exists. By Theorem 2, the right maximal solution $s(t)$ passing through (a, x_0) either exists on $[a, b]$ or tends to the boundary of D as $t \to c$, where $c \in (a, b]$, and then exists on $[a, c)$. But then $s(t) \in S$ on $[a, c)$, so as $t \to c$, any limit points of $s(t)$ lie in S , a compact subset of D , and so cannot be on the boundary of D .

So $s(t)$ exists and $y(t) > s(t) > z(t)$, on $[a, b]$.

Theorem 5 can be much improved if $f(t, y)$ is sufficiently smooth on D to ensure uniqueness; that is if, for example, the one-sided Lipschitz condition of the following Theorem 7 applies on D . In this case the strong inequalities can be replaced by weak inequalities:

THEOREM 6. *Suppose the solution of the initial value problem* $y' = f(t, y)$ *,* $y(t_0) = y_0$ *,* $(t_0, y_0) \in D$ *is unique whenever it exists in* D *when* $t \ge t_0$ *for all initial values* (t_0, y_0) *in* D *,* *and suppose* $y(t)$, $z(t)$ *are continuous functions, and are solutions of the differential inequalities* $D^{-}y \geq f(t, y)$, $D_{-}z \geq f(t, z)$ *respectively, where* $(t, y(t))$, $(t, z(t)) \in D$.

Suppose that for some point a , $z(a) \leq y(a)$, *and* x_0 *is any vector for which* $z(a) \leq x_0 \leq y(a)$. *Then the initial value problem*

$$x' = f(t, x) ,$$
$$x(a) = x_0$$

has a solution $x(t)$ *which is defined and satisfies the inequalities* $z(t) \leq x(t) \leq y(t)$ *in* D .

Proof. From Theorem 2, $x(t)$ exists in some interval $[a, c)$ and tends to the boundary of D as $t \to c$. But if $(y(c), c)$, $(z(c), c)$ are in D then $u : y(c) \geq u \geq z(c)$ is a compact set contained in an open set; it is therefore distant d from the boundary of the open set, for some positive scalar d .

But by Theorem 3, $y(t) \geq x(t) \geq z(t)$ in $[a, c)$, that is, $x(t) \in K(t)$ if $K(t)$ is the function defined in Theorem 5. But $K(t)$ is continuous, so

$$S = \bigcup_{t \in [a,c), x \in K(t)} (x, t)$$

is a compact set, and so contains $\lim_{t \to c} (x(t), t)$. Therefore $\lim_{t \to c} x(t) \in K(c)$, and is not on the boundary of D . So $x(t)$ exists everywhere in D .

Application to establishing one-sided uniqueness theorem

As an application of Theorem 3, a demonstration is given of the use of a one-sided Lipschitz condition, suggested by W.A. Coppel.

THEOREM 7. *Let* $f(t, x)$, $g(t, x)$ *be continuous functions from* $R \times R^n$ *to* R^n , *where the domain of* f *is an open set* D *in* $R \times R^n$, *and the domain of* g *includes all points* $(t, x-y)$, *where* (t, x), $(t, y) \in D$.

Suppose g *is of type* K_+ *on its domain, and*

$$f(t, x) - f(t, y) \le g(t, x-y) \quad \forall (t, x), (t, y) \in D .$$

Let $y_1(t), y_2(t)$ *be solutions of*

$$y_1' \ge f(t, y_1) ,$$

$$y_2' \le f(t, y_2) ,$$

respectively, with $(t_0, y_1(t_0))$ *and* $(t_0, y_2(t_0)) \in D$. *Then* $y_2(t) - y_1(t) \le w(t)$ *if* $t \ge t_0$ *, where* $w(t)$ *is a right maximal solution of* $w' = g(t, w)$, $w(t_0) = y_2(t_0) - y_1(t_0)$.

Proof. Let $y(t) = y_2(t) - y_1(t)$. Then

$$\begin{aligned} y'(t) &\le f(t, y_2) - f(t, y_1) \\ &\le g(t, y) . \end{aligned}$$

Therefore $y(t) \le w(t)$ for $t \ge t_0$, by Theorem 3.

COROLLARY. *If* f *is of type* K_+ *, and* $w \equiv 0$ *a solution of* $w' = g(t, w)$ *for* $t \ge t_0$ *there is at most one solution to the right of* t_0 *of the initial value problem* $y' = f(t, y), y(t_0) = y_0$. *For if there are two,* $y_1(t)$ *and* $y_2(t)$ *, then by the theorem just proved* $y_1(t) - y_2(t) \le 0$ *and* $y_1(t) - y_2(t) \ge 0$ *if* $t \ge t_0$.

Suppose $g(t, x) = \alpha x$ for some constant $\alpha > 0$. Then $g(t, x)$ is of type K_+ everywhere, and $x \equiv 0$ is a solution of $x' = g(t, x)$. This special case corresponds to the Lipschitz condition as normally defined. If $f(t, x) - f(t, y) \le \alpha(x-y)$ for all $(t, x), (t, y)$ in some domain D then if $(t_0, x_0) \in D$ there is at most one solution $y(t)$ of (1) with $y(t_0) = x_0$, if $t > t_0$.

Examples of various cones and orderings

1) $C = C^* = \{\{x_i\} : x_i \ge 0, i = 1, \ldots, n\}$ in R^n .

This generates the usual partial ordering on a vector space, and specialises to the differential inequalities previously dealt with. (Coppel [2], Szarski [1], Walter [1].)

2) $C = \{(x, y, z) : x \geq 0, z \geq 0, xz \geq y^2\}$.

This generates a partial order equivalent to that of the 2×2 symmetric matrices ordered by the positive definiteness relation. Then $C^* = \{(x, y, z) : x \geq 0, z \geq 0, 4xz \geq y^2\}$ and C^* is generated by the set of vectors $(x^2, 2xy, y^2)$.

EXAMPLE 3. Here the cone C is the circular Lorentz cone: a vector $\mathbf{x} \in C$ if $x_1 \geq 0$ and $x_1^2 \geq x_2^2 + \ldots + x_n^2$. Put another way, C is the set of all vectors $(\alpha, \mathbf{a})$ where $\alpha \geq 0$, $\mathbf{a}$ is a $(n-1)$-vector, and $|\mathbf{a}| \leq \alpha$. Then C is self-dual. For if $(\alpha, \mathbf{a})$, $(\beta, \mathbf{b})$ are two vectors in C , then

$$(\alpha, \mathbf{a}).(\beta, \mathbf{b}) = \alpha\beta + \mathbf{a}.\mathbf{b} \geq \alpha\beta - |\mathbf{a}||\mathbf{b}| \geq 0 .$$

Conversely, if $(\beta, \mathbf{b})$ is any vector in C^* , then $\beta = (\beta, \mathbf{b}).(1, 0) \geq 0$ since $(1, 0) \in C$. And $(|\mathbf{b}|, -\mathbf{b}) \in C$, so $\beta|\mathbf{b}| - \mathbf{b}.\mathbf{b} \geq 0$; that is, $\beta \geq |\mathbf{b}|$ (even if $|\mathbf{b}| = 0$) so $(\beta, \mathbf{b}) \in C$. Therefore $C^* = C$.

Let $\mathbf{a}, \mathbf{b}$ be arbitrary n-vectors, α any scalar, and $\mathbf{s}_i$ the column vectors of an anti symmetric $n \times n$ matrix S . Let $\{\ \}$ denote the Lorentz type scalar product, that is,

$$\{a, b\} = a_1b_1 - a_2b_2 - \ldots - a_nb_n ,$$

and let $\mathbf{f}(\mathbf{x})$ be the n-vector valued function whose components are

$$f_i(\mathbf{x}) = a_i\{\mathbf{x}, \mathbf{x}\} - 2x_i\{\mathbf{a}, \mathbf{x}\} + \{\mathbf{s}_i, \mathbf{x}\} + \alpha x_i + b_i . \qquad (4)$$

If two vectors $\mathbf{x}, \mathbf{y}$ in C have the property that $\mathbf{x}.\mathbf{y} = 0$, then $\mathbf{x}$ must be a multiple of $(1, \boldsymbol{\ell})$ and $\mathbf{y}$ of $(1, -\boldsymbol{\ell})$ where $\boldsymbol{\ell}$ is a unit $(n-1)$-vector.

With this observation it is easy to verify that $\mathbf{f}(\mathbf{x})$ is both of type K_+ and K_- with respect to C .

So solutions of the equation

$$\mathbf{x}' = \mathbf{f}(t, \mathbf{x}) , \qquad (5)$$

where the coefficients a, b, α, S are continuous functions of t, have the property that any ordering of solutions at a point with respect to the Lorentz cone is preserved as t changes, while the respective solutions continue to exsit.

In the four-dimensional space-time of relativity theory, if two points, or "events" x and y are ordered $(\mathrm{x} \geq \mathrm{y})$ with respect to C, then x is "attainable" from y, or y is "observable" from x. Any useful transformation of space-time will need to preserve these relationships at all points under consideration.

The usual Lorentz transformations can be derived from special cases of (5), with $\mathrm{a}(t) \equiv \mathrm{b}(t) \equiv \alpha(t) = 0$, and using a simpler version of $S(t)$.

Example 3 is of particular interest because it is closely related to the symmetric matrix ordering to be developed in the following chapters. In fact if $n = 3$, and the orientation of the co-ordinate axes is changed, the 2×2 matrix ordering is obtained. More is said about this example in the appendix to Chapter 3.

Notes

There have been many publications recently dealing with differential inequalities, both in finite-dimensional and more general vector spaces. Szarski [*1*], Walter [*1*], Lakshmikantham and Leela [*1*], Coppel [*2*] give expositions which, for the componentwise vector ordering, go much further than is necessary for our purposes.

Some authors (Mlak [*1*], Cohen and Lees [*1*], Edmunds [*1*]) have extended some results about differential inequalities to Hilbert or Banach spaces. In Mlak's paper, the ordering used is again componentwise, and in the other papers the results are obtained by comparison with a finite-dimensional system of the usual type.

Our type K condition is more often referred to as monotonicity or quasimonotonicity, the latter name being introduced by Walter. For componentwise ordering, it was mentioned by Müller (1926), and used by Kamke (1932). Geometrically, its significance is that if $u(t)$, $v(t)$ are solutions of $y' = f(t, y)$, $x(t) = u(t) - v(t)$, and $x(t)$ comes to the edge of the positive cone, then the derivative of $x(t)$ is not in a direction leading out of the cone, that is, it is either parallel to the surface or towards the interior. This is true of our definition for more general cones, and the consequence is that $x(t)$ must remain within the cone, if its initial value is in the cone. It is likely that this result could be extended to convex sets generally, not just cones; but then to express the corresponding type K condition analytically is very difficult.

Coppel [3] has obtained results equivalent to our Theorems 1-3 and 6 for symmetric solutions of the matrix Riccati equation. In this more simple case, direct methods are also available, and implicitly or explicitly, the preservation of ordering of solutions of the Riccati equation has been known and used for some time. Indeed, the property is closely related to Stürm-Liouville comparison theory for Hamiltonian systems.

Yorke [7] investigates whether solutions of a differential equation can be constrained to lie within certain sets; our approach deals with differences of solutions, but some of the ideas are similar.

CHAPTER 2

MATRIX EQUATIONS AND A PROPERTY UNIQUE TO THE RICCATI EQUATION

Introduction

This chapter gives details of the application of the general inequalities of Chapter 1 to symmetric matrix systems of differential equations and specifically to the matrix Riccati equation. Writing a system of differential equations in a matrix form, symmetric or otherwise, adds nothing new except convenience and an awareness of the possibilities of various manipulations that may not have been otherwise evident. The real difference is in the choice of ordering system; it is convenient to use the positive definiteness-ordering of symmetric matrices, where $A \geq B$ if $\xi^*(A-B)\xi \geq 0$ for all vectors ξ .

The first task is to verify that this ordering does fulfil the requirements of Chapter 1, so the results of that chapter can be applied. It is then shown that the matrix Riccati equation is both of type K_+ and K_- . In this case any order relation that may exist between two solutions at some point is preserved as this point moves in either direction, so that it is meaningful to speak of one solution being greater than another without specifying a point. Put another way, the solutions existing on an interval are ordered according to their values at any point on the interval. Extensive use of this property is made later.

Finally, it is proved that with respect to this latter property the matrix Riccati equation is unique among matrix systems. This rather surprising result is analogous to the result for vector systems ordered by the usual partial ordering, that a function $f(x)$ both of type K_+ and K_- must be of the form $(f_1(x_1), f_2(x_2) \ldots f_n(x_n))$ which leads to a trivial system (Szarski [1]).

The matrix ordering

The set of $n \times n$ symmetric matrices form a vector space M_n

of dimension $\frac{1}{2}n(n+1)$. In this space, we define the scalar product of two elements $A = \{a_{ij}\}$, $B = \{b_{ij}\}$ to be

$$(A, B) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij} = \mathrm{Tr}(AB) \ .$$

The set of non-negative definite symmetric matrices forms a closed cone C in M_u .

LEMMA. *C is self-dual, that is, $C = C^*$.*

Proof. a) If $A \in C$, $B \in C$ then for some matrix $T = \{t_{ij}\}$, $B = T^*T$. Then

$$\begin{aligned}(A, B) &= \mathrm{Tr}(AB) \\ &= \mathrm{Tr}(AT^*T) \\ &= \mathrm{Tr}(TAT^*) \geq 0 \ .\end{aligned}$$

This is true for arbitrary A and B so $B \in C^*$ and $C \subset C^*$.

b) Let A be any matrix in C^* . Let $X = \{x_i x_j\} \in C$ for any n-tuplet $\{x_i\}$, so

$$\begin{aligned}(A, X) &= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \\ &= x^*Ax \geq 0\end{aligned}$$

where $x = \{x_i\}$. Therefore, $A \in C$ and $C^* \subset C$. Therefore, $C = C^*$.

In this section we will exclude the trivial case where $n = 1$ and always assume $n \geq 2$.

In the case of the present cone C , and also in most other applications of the theory of Chapter 1, the type K condition can be simplified with the aid of special knowledge about when it is possible to have vectors x in C and ξ in C^* for which $(x, \xi) = 0$.

The type K condition needed for the theorems of Section I is:

DEFINITION I. *A symmetric matrix function $F(X)$ of the symmetric matrix variable X , defined on a domain D is of type*

$K_+\{K_-\}$ *on* D *if, whenever* $W_1 \in D$, $W_2 \in D$, $W_1 - W_2 \geq 0$ *and for some matrix* $A \in C$, $(W_1-W_2, A) = 0$, *then* $(F(W_1)-F(W_2), A) \geq 0$ $\{\leq 0\}$.

The simpler version is:

DEFINITION II. *If* $F(X)$ *is the function of Definition* I, *it is of type* K_+ $\{K_-\}$ *if, whenever* $W_1 \in D$, $W_2 \in D$, $W_1 - W_2 \geq 0$ *and for some vector* $\xi \in R^n$, $W_1\xi = W_2\xi$, *then*

$$\xi^*(F(W_1)-F(W_2))\xi \geq 0 \quad \{\leq 0\} . \tag{1}$$

If $F(X)$ is of type K_+ according to I, then it is also according to II, since the matrix $\{\xi_i\xi_j\}$ is in C .

If $F(X)$ is of type K_+ according to II, and there are matrices $W_1 \in D$, $W_2 \in D$, $A \in C$: $(W_1-W_2) \in C$ and $(W_1-W_2, A) = 0$ then there is a matrix T : $A = T^*T$, where $T = \{t_{ij}\}$ and

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (W_1-W_2)_{ij} t_{ki} t_{kj} = (W_1-W_2 \ , A) = 0 .$$

Let $\mathbf{t}_k = \{t_{ki}\}$, then $\sum_{k=1}^{n} \mathbf{t}_k^*(W_1-W_2)\mathbf{t}_k = 0$.

Since $W_1 \geq W_2$, for each k , $\mathbf{t}_k^*(W_1-W_2)\mathbf{t}_k = 0$, so $(W_1-W_2)\mathbf{t}_k = 0$. Therefore $\mathbf{t}_k^*(F(W_1)-F(W_2))\mathbf{t}_k \geq 0$ and summing over k , $(F(W_1)-F(W_2), A) \geq 0$.

So both definitions of type K_+ (and K_-) are equivalent. Therefore the theorems of Section I can be applied using either definition.

The matrix Riccati operator $R[W]$ is defined by

$$R[W] = W' + A(t) + B(t)W + WB^*(t) + WC(t)W \tag{2}$$

where W, $A(t)$, $B(t)$, $C(t)$ are $n \times n$ matrices, continuous functions of t in an interval I , and $A(t)$, $C(t)$ are symmetric on I .

Then the Riccati equation is

$$R[W] = 0 . \tag{3}$$

Applying Definition II above, it is clear that (3) is both of type K_+ and K_- everywhere. For if

$$F(t, W) = A(t) + B(t)W + WB^*(t) + WC(t)W ,$$

and $W_1\xi = W_2\xi = \eta$, then

$$\begin{aligned}\xi^*F(t, W_1)\xi &= \xi^*A(t)\xi + \xi^*B(t)\eta + \eta^*B^*(t)\xi + \eta^*C(t)\eta \\ &= \xi^*F(t, W_2)\xi .\end{aligned}$$

The following theorem conveniently summarises Theorems 3 and 6 of Chapter 1 in this matrix context:

THEOREM 1. *Let* W_1, W_2 *be any* $n \times n$ *symmetric matrices for which* $W_1 \geq W_2$. *Let* $W_1(t)$, $W_2(t)$ *be solutions of the inequalities* $\operatorname{sgn}(t-t_0)R[W_1] \geq 0$, $\operatorname{sgn}(t-t_0)R[W_2] \leq 0$ *respectively existing on some interval* $[a, b]$ *containing* t_0 *, with* $W_1(t_0) = W_1$ *,* $W_2(t_0) = W_2$. *Then* $W_1(t) \geq W_2(t)$ *on* $[a, b]$. *And if* $W(t)$ *is a solution of* (3) *with* $W_1 \geq W(t_0) \geq W_2$ *, then* $W(t)$ *exists on* $[a, b]$ *and* $W_1(t) \geq W(t) \geq W_2(t)$.

In particular, if $R[W_1] = R[W_2] = 0$ *on* $[a, b]$ *, and* $W_1(t) \geq W_2(t)$ *at some point* t_0 *, then* $W_1(t) \geq W_2(t)$ *on* $[a, b]$.

Uniqueness of the Riccati equation

Theorem 1 is a strong result, and applies only to the Riccati equation (if $n \geq 2$). The converse result is:

THEOREM 2. *If* $F(t, Y)$ *is an* M_n*-valued function defined and continuous in* Y *for each* t *on some domain* D *in* $R \times M_n$ *,* $n \geq 2$ *, and if it is true that whenever* W_1, W_2 *are two matrices,* t_0 *a point for which* (t_0, W_1) *and* (t_0, W_2) *are in* D *, and* $W_1 \geq W_2$ *, there are two solutions* $W_1(t)$, $W_2(t)$ *of* $W' = F(t, W)$

with $W_1(t_0) = W_1$, $W_2(t_0) = W_2$ *for which* $W_1(t) \geq W_2(t)$ *in some neighbourhood of* t_0 , *then it follows that* $Y' - F(t, Y)$ *is of the form* $R[Y]$ *for some coefficient matrices defined for* t *in* D .

The proof of this theorem occupies the rest of the chapter. It will be assumed that D includes all M_n for each t . This avoids having to make frequent provisos about the domain in an already complicated proof; it is not an important restriction because the proof is local in character.

The hypothesis of the theorem is just what is needed to apply Corollary 2 of Theorem 4, Chapter 1, which ensures that $F(t, Y)$ is both of type K_+ and K_- for each t . This said, no further interaction between t and Y occurs, so mention of t will be suppressed.

So if $W_1 \geq W_2$ and $W_1\xi = W_2\xi$ then $\xi^*F(W_1)\xi = \xi^*F(W_2)\xi$. If W_1 and W_2 are any two symmetric matrices, with $W_1\xi = W_2\xi$, there is another symmetric matrix W_3 for which $W_3\xi = W_1\xi$, $W_3 \geq W_1$, $W_3 \geq W_2$. Then

$$\xi^*F(W_1)\xi = \xi^*F(W_3)\xi = \xi^*F(W_2)\xi .$$

Therefore, $\xi^*F(W)\xi$ is a function of ξ and $W\xi$ only. So for some function g on R^{2n} ,

$$\xi^*F(W)\xi = g(\xi, \eta) \quad \text{where} \quad \eta = W\xi . \tag{4}$$

Then equation (4) is restrictive enough to ensure that, except in the trivial case where $R^n = R^1$, $W' - F(t, W)$ has the form $R[W]$ of the Riccati equation, for some appropriate set of coefficients, and $g(\xi, \eta)$ is a quadratic form in ξ and η .

Since this is not readily apparent at first sight, and since the reason for it is not much illuminated by the rather complex proof, the following lemma is given to show a simple proof when g is assumed smooth when its domain has been extended to all of R^{2n} .

LEMMA. *If* $g(x, y)$ *is a function defined and* C^2 *in some neighbourhood of* $(0, 0)$ *then* $g(x, y) = x^*Ax + x^*By + y^*B^*x + y^*Cy$

for some coefficients A, B, C *if* (4) *holds.*

Proof. For any scalar t,

$$g(tx,\ ty) = t^2x^*F(W)x$$
$$= t^2g(x,\ y)\ .$$

So $g(0,\ 0) = 0$, and $g(-x,\ -y) = g(x,\ y)$. Then

$$2g(x,\ y) = \frac{1}{t^2}[g(tx,\ ty)+g(-tx,\ -ty)-2g(0,\ 0)]$$
$$\to 2[x^*Ax+x^*By+y^*B^*x+y^*Cy] \quad \text{as} \quad t \to 0$$

where

$$A_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} g(0,\ 0)\ , \quad B_{ij} = \frac{\partial^2}{\partial x_i \partial y_j} g(0,\ 0)\ , \quad C = \frac{\partial^2}{\partial y_i \partial y_j} g(0,\ 0)\ .$$

Therefore, $g(x,\ y) = x^*Ax + x^*By + y^*B^*x + y^*Cy$. QED

The next, and longest, step in the proof of Theorem 2, is to show that each coefficient of $F(W)$ is quadratic in some of the coefficients of W .

Let e_i be the unit vector whose i-th component is 1 . Then $F_{ii}(W) = e_i^*F(W)e_i = g(e_i,\ We_i)$. So $F_{ii}(W)$ is a function only of the coefficients W_{ij} , $j = 1, \ldots, n$. And

$$2F_{ij}(W) = (e_i+e_j)^*F(W)(e_i+e_j) - F_{ii}(W) - F_{jj}(W)$$
$$= g(e_i+e_j,\ We_i+We_j) - g(e_i,\ We_i) - g(e_j,\ We_j)\ .$$

So $F_{ij}(W)$ is a function of W_{ik} and W_{jk} only, $k = 1 \ldots n$.

The problem is now artificially restricted to a 2 × 2 problem, as follows. For arbitrary i, j , $i \neq j$, let $W_1 = W_{ii}$, $W_2 = W_{ij}$, $W_3 = W_{jj}$, and assume during what follows that all other coefficients of W remain fixed. Let

$$F_1(W) = F_{ii}(W)\ , \quad F_2(W) = F_{ij}(W)\ , \quad F_3(W) = F_{jj}(W)\ .$$

Suppressing constant coefficients, and letting α be any constant, (4) implies

$$\alpha^2F_1(W_1,\ W_2)\ +\ 2\alpha F_2(W_1,\ W_2,\ W_3)\ +\ F_3(W_2,\ W_3)\ =$$
$$=\ g(\alpha,\ 1,\ \alpha W_1+W_2,\ \alpha W_2+W_3)\ .\quad (5).$$

Neither side of (5) is affected by the change

$$W_1 \to W_1 + \varepsilon\ ,$$
$$W_2 \to W_2 - \alpha\varepsilon\ ,$$
$$W_3 \to W_3 + \alpha^2\varepsilon\ ,$$

for any ε . Therefore,

$$\alpha^2F_1(W_1+\varepsilon,\ W_2-\alpha\varepsilon)\ +\ 2\alpha F_2\left(W_1+\varepsilon,\ W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)\ +\ F_3\left(W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)$$
$$=\ \alpha^2F_1(W_1,\ W_2)\ +\ 2\alpha F_2(W_1,\ W_2,\ W_3)\ +\ F_3(W_2,\ W_3)\ .\quad (6)$$

This is the basic equation to be manipulated; it is rewritten by first replacing α by $-\alpha$ and ε by $-\varepsilon$, then $(W_1,\ W_2,\ W_3)$ by $\left(W_1+\varepsilon,\ W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)$, so:

$$\alpha^2F_1(W_1,W_2-2\alpha\varepsilon)\ -\ 2\alpha F_2(W_1,\ W_2-2\alpha\varepsilon,\ W_3)\ +\ F_3(W_2-2\alpha\varepsilon,\ W_3)$$
$$=\ \alpha^2F_1(W_1+\varepsilon,\ W_2-\alpha\varepsilon)\ -\ 2\alpha F_2\left(W_1+\varepsilon,\ W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)\ +$$
$$+\ F_3\left(W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)\ .\quad (7)$$

Adding (6) and (7):

$$\alpha^2\left[F_1(W_1,\ W_2-2\alpha\varepsilon)-F_1(W_1,\ W_2)\right]\ +\ F_3(W_2-2\alpha\varepsilon,\ W_3)\ -\ F_3(W_2,\ W_3)$$
$$+\ 2\alpha\left[2F_2\left(W_1+\varepsilon,\ W_2-\alpha\varepsilon,\ W_3+\alpha^2\varepsilon\right)-F_2(W_1,\ W_2-2\alpha\varepsilon,\ W_3)-F_2(W_1,\ W_2,\ W_3)\right]$$
$$=\ 0\ .\quad (8)$$

Dividing by α and letting $\alpha \to 0$:

$$\lim_{\alpha\to 0}\ \frac{F_3(W_2-2\alpha\varepsilon,W_3)-F_3(W_2,W_3)}{2\alpha\varepsilon}$$
$$=\ -\ \frac{2}{\varepsilon}\left[F_2(W_1+\varepsilon,\ W_2,\ W_3)-F_2(W_1,\ W_2,\ W_3)\right]\ .\quad (9)$$

Therefore, $\dfrac{\partial F_3}{\partial W_2}\,(W)$ exists for all $\alpha,\ \varepsilon$. We abbreviate $\dfrac{\partial F_i}{\partial W_j}$ by

F_{ij} , $i, j = 1, 2, 3$. Then

$$F_2(W_1+\varepsilon, W_2, W_3) = F_2(W_1, W_2, W_3) + \tfrac{1}{2}\varepsilon F_{32}(W_2, W_3) . \tag{10}$$

So F_2 is a linear function of W_1 for fixed W_2, W_3 .

In (8) let $\alpha \to \infty$ and $\varepsilon = \dfrac{t}{\alpha^2}$ for some constant t . Then dividing by α :

$$\lim_{\alpha\varepsilon\to 0}\left[\frac{t}{2\alpha\varepsilon}\left(F_1(W_1, W_2-2\alpha\varepsilon)-F_1(W_1, W_2)\right)\right] + 2\left[F_2(W_1, W_2, W_3+t)-F_2(W_1, W_2, W_3)\right] = 0 . \tag{11}$$

Therefore, $\dfrac{\partial F_1}{\partial W_2} = F_{12}(W_1, W_2)$ exists and

$$tF_{12}(W) = 2\left[F_2(W_1, W_2, W_3+t)-F_2(W_1, W_2, W_3)\right] , \tag{12}$$

that is, F_2 is also a linear function of W_3 for fixed W_1, W_2 .

Rewriting (6), replacing α by $-\alpha$, (W_1, W_2, W_3) by $\left(W_1+\varepsilon, W_2-\alpha\varepsilon, W_3+\alpha^2\varepsilon\right)$, then

$$\begin{aligned}\alpha^2F_1(W_1+2\varepsilon, W_2) &- 2\alpha F_2\left(W_1+2\varepsilon, W_2, W_3+2\alpha^2\varepsilon\right) + F_3\left(W_2, W_3+2\alpha^2\varepsilon\right) \\ &= \alpha^2F_1(W_1+\varepsilon, W_2-\alpha\varepsilon) - 2\alpha F_2\left(W_1+\varepsilon, W_2-\alpha\varepsilon, W_3+\alpha^2\varepsilon\right) + F_3\left(W_2-\alpha\varepsilon, W_3+\alpha^2\varepsilon\right)\end{aligned}$$

and adding (6) to this equation,

$$\begin{aligned}&\alpha^2\left[F_1(W_1+2\varepsilon, W_2)-F_1(W_1, W_2)\right] + F_3(W_2, W_3+2\alpha^2\varepsilon) - F_3(W_2, W_3) \\ &+ 2\alpha\left[2F_2\left(W_1+\varepsilon, W_2-\alpha\varepsilon, W_3+\alpha^2\varepsilon\right)-F_2\left(W_1+2\varepsilon, W_2, W_3+2\alpha^2\varepsilon\right)-F_2(W_1, W_2, W_3)\right] \\ &= 0 .\end{aligned} \tag{13}$$

F_2 is a linear function of W_1 , so $F_1(W_1+2\varepsilon, W_2) - F_1(W_1, W_2)$ is also linear in W_1 , from (13). Suppressing W_2 for the time being,

$$\begin{aligned}\left[F_1(W_1+3\varepsilon)-F_1(W_1+2\varepsilon)\right] &- \left[F_1(W_1+2\varepsilon)-F_1(W_1+\varepsilon)\right] \\ &= \left[F_1(W_1+2\varepsilon)-F_1(W_1+\varepsilon)\right] - \left[F_1(W_1+\varepsilon)-F_1(W_1)\right] ,\end{aligned}$$

that is,

$$F_1(W_1+3\varepsilon) - 3F_1(W_1+2\varepsilon) + 3F_1(W_1+\varepsilon) - F_1(W_1) = 0 \ . \qquad (14)$$

Given any three values of F_1, say $F_1(1)$, $F_1(0)$, $F_1(-1)$, then (14) can be used to determine values at all integer points, and the consequent equation,

$$8F_1(W_1+\varepsilon) = 3F_1(W_1+2\varepsilon) + 6F_1(W_1) - F_1(W_1-2\varepsilon) \ ,$$

all $(m+\frac{1}{2})$ values of F_1 (m any integer) and so on, giving values at any argument of the form $\frac{p}{2^q}$, p, q any integers. The latter points are dense in the continuum, so F_1 is determined by (14), given any three of its values. But there is exactly one quadratic solution of (14) with those three values. So F_1 must be quadratic in W_1.

Similarly F_3 is a quadratic function of W_3.

We return now to n dimensions, and the original notation for coefficients of F and W. Whenever a vector x has no zero components, then for any $y \in R^n$,

$$g(x, y) = x^*F(W)x$$

where $W_{ii} = \frac{y_i}{x_i}$, $i = 1 \ldots n$, and $W_{ij} = 0$ if $i \neq j$. Then

$$g(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j F_{ij}(W) \ .$$

In this sum, in the cases when $i = j$, F_{ij} is a quadratic function of $W_{ii} = \frac{y_i}{x_i}$ and independent of all other variables, so $x_i^2 F_{ii}(W)$ is a homogeneous quadratic form in x_i, y_i.

And if $i \neq j$, F_{ij} is a function of W_{ii}, W_{ij} only, and is linear in each taken independently. So again $x_i x_j F_{ij}\left(\frac{y_i}{x_i}, \frac{y_j}{x_j}\right)$ is a homogeneous quadratic form in x_i, x_j, y_i, y_j .

So $g(x, y)$ is a homogeneous quadratic function in x and y , unless some coefficient of x is zero. But $g(x, y)$ is also continuous, (except when $x = 0$) so is a homogeneous quadratic form everywhere. Although it is not defined by (4) when $x = 0$, the domain of definition can be extended to include such points. So

$$g(x, y) = (x^*, y^*)\begin{pmatrix} A & B \\ B^* & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

where A, B, C are $n \times n$ matrices, $A = A^*$, $C = C^*$. Therefore

$$\begin{aligned} g(x, Wx) &= x^*Ax + x^*BWx + x^*WB^*x + x^*WCWx \\ &= x^*F(W)x \ . \end{aligned}$$

So $F(W) = A + BW + WB^* + WCW$.

To get this result, a fixed value of t was used. The coefficients A, B, C will generally be functions of t . They need not be continuous, but the order-preserving property, as stated, will impose some limitations on their behaviour. If $F(t, W)$ is assumed continuous in t , then $A(t), B(t), C(t)$ are continuous also. This can be shown by considering special W values (for example, $A(t) = F(t, 0)$).

Notes for Chapter 2

For Theorem 1, derived by a similar method whose application is there restricted to matrix Riccati equations, see Coppel [3]. Some of the conclusions are used, more or less explicitly, in many other papers; see, for example, Reid [9] and [15].

Reid's paper [15] also contains a general non-linear matrix equation which has order-preserving properties. It involves, however, a monotone function of a matrix variable, which is rather a strong requirement. The order-preserving properties are restricted to non-negative solutions.

CHAPTER 3

MAXIMAL SOLUTIONS OF THE MATRIX RICCATI EQUATION AND DISCONJUGACY

Introduction

In this chapter two basic systems of differential equations are considered. The matrix Riccati equation is defined

$$R[W] \equiv W' + A(t) + WB(t) + B^*(t)W + WC(t)W = 0 \tag{1}$$

and the corresponding Hamiltonian system is

$$\begin{aligned} Y' &= B(t)Y + C(t)Z \ , \\ Z' &= -A(t)Y - B^*(t)Z \ . \end{aligned} \tag{2}$$

In each case the coefficient $n \times n$ matrices $A(t)$, $B(t)$, $C(t)$ are defined and continuous in an interval I , $A(t)$ and $C(t)$ are symmetric in I , and, unless otherwise indicated $C(t) \geq 0$ is assumed on I .

A matrix function $W(t)$ will be called a solution of (1) only if it is symmetric. A solution $\langle Y(t), Z(t) \rangle$ of (2) will be a pair of $n \times n$ matrices, differentiable on I and satisfying (2), with the following properties:

a) Isotropy: $Y^*(t)Z(t) = Z^*(t)Y(t)$ on I .

b) Non-degeneracy: If $Y(t)\xi = Z(t)\xi = 0$, then $\xi = 0$, where t is any point on I .

There is an associated vector system

$$\begin{aligned} y' &= B(t)y + C(t)z \ , \\ z' &= -A(t)y - B^*(t)z \ , \end{aligned} \tag{2a}$$

where y, z are n-dimensional vectors. A solution $\langle y(t), z(t) \rangle$ of (2a) is always of the form $\langle Y(t)\xi, Z(t)\xi \rangle$ for some solution $\langle Y, Z \rangle$ of (2), and some constant vector ξ .

It is easy to verify that the restrictions of symmetry (for (1)) and isotropy and non-degeneracy are met wherever the appropriate solutions exist, provided the restrictions are met at a single point.

Solutions of (1) and (2) are closely related. If $\langle Y(t), Z(t) \rangle$ is a solution of (2) with $Y(t)$ invertible at some point t_0 , then $W(t) = Z(t)Y^{-1}(t)$ exists in some neighbourhood of t_0 and is a solution of (1). Conversely if $W(t)$ is a solution of (1) on I, and $Y(t)$ is a solution of

$$Y' = [B(t)+C(t)W(t)]Y \tag{3}$$

invertible at some point in I (and so throughout I), then $\langle Y(t), W(t)Y(t)\rangle$ is a solution of (2) throughout I. Invertibility of $Y(t)$ ensures non-degeneracy of $\langle Y(t), W(t)Y(t)\rangle$.

The following basic properties of (2) can be verified by direct differentiation: if $\langle Y_1(t), Z_1(t)\rangle$, $\langle Y_2(t), Z_2(t)\rangle$ are solutions of (2) (not necessarily isotropic or non-degenerate) on I, then

(i) $Z_2^*(t)Y_1(t) - Y_2^*(t)Z_1(t) = N$, a constant matrix on I; (4)

(ii) $\{Y_1^*(t)Z_1(t)\}' = Z_1^*(t)C(t)Z_1(t) - Y_1^*(t)A(t)Y_1(t)$. (5)

The main results of this chapter fall into two classes. In Theorem 1, taking the existence of a solution of (1) on an interval as the basic condition, called $[R]$, a "principal" solution of (2) is established, corresponding to a "maximal" or "minimal" solution of (1). It does not matter whether the interval in question is compact, half-open or open, nor whether (1) obeys any controllability or normality type conditions.

Later, culminating in Theorem 4, we relate $[R]$ to disconjugacy, giving a necessary and sufficient condition. This approach is interesting because it avoids variational arguments and the necessity of establishing results firstly for compact intervals, and also because controllability conditions appear in a secondary role, merely conferring extra properties on the principal solutions, but not affecting the fundamental behaviour of the systems.

Most of the proofs are given for left-hand or right-hand endpoints only. The symmetry of the results is usually obvious and sometimes assumed for later proofs. By way of summary, dual versions of the main results are given at the end of the chapter.

Regularity conditions and their interrelations

In connection with (1) and (2), consequences of various of the following conditions will be of interest:

$[R]$: *a solution of* (1) *exists on* I , *or, equivalently, a solution* $\langle Y(t),\ Z(t)\rangle$ *of* (2) *exists, and* $Y(t)$ *is invertible, on* I .

$[WD]$ *(weak) disconjugacy: if* $\langle Y(t),\ Z(t)\rangle$ *is a solution of* (2) *in* I , *and* $[a,\ b]$ *is any compact interval in* I , *then* $Y(a)\xi = Y(b)\xi = 0$ *only if* $Y(t)\xi = 0$ *in* $[a,\ b]$.

$[D]$ *(strong) disconjugacy: if* $\langle Y(t),\ Z(t)\rangle$ *is a solution of* (2) *in* I , *and* $[a,\ b]$ *is any compact interval in* I , *then* $Y(a)\xi = Y(b)\xi = 0$ *only if* $\xi = 0$.

$P(J)$: *If* $\Omega(t)$ *is a fundamental matrix of the equation* $Z' = -B^*(t)Z$ *on* J , *then for a vector* ξ , $C(t)\Omega(t)\xi \equiv 0$ *on* J *iff* $\xi = 0$.

$M_+(b)$: $P\big((b,\ c]\big)$ *holds for all* c *in some right neighbourhood of* b .

$M_-(b)$: $P\big([c,\ b)\big)$ *holds for all* c *in some left neighbourhood of* b .

$[C]$: $P(J)$ *holds for all sub-intervals* J *of* I .

The following relations hold between the various conditions.

a) On any interval $[R] \Rightarrow [WD]$. On compact intervals $[WD] \Rightarrow [R]$. (Theorem 2.) If $[WD]$ holds on a half-open subinterval, $[R]$ holds on a half-open sub-interval having the same open end-point (Theorem 3). For an extended concept of disconjugacy, and clarification of this position, see Theorem 4.

b) On a half-open interval $(a,\ b]$, $M_+(a)$ and $[R]$ are equivalent to the existence of a maximal solution of (1) on $(a,\ b]$. And on $[b,\ a)$, $M_-(a)$ and $[R]$ are equivalent to the existence of a minimal solution of (1) on $[b,\ a)$.

c) $[C]$ and $[R] \Longleftrightarrow [D]$ for open and compact intervals. For any interval, $[C]$ and $[R] \Rightarrow [D]$, $[D] \Rightarrow [C]$ and on any open or compact subinterval $[D] \Rightarrow [R]$.

Remarks on maximal and principal solutions

The context (and results) of this chapter have been indicated without much explanation or motivation. The three levels of conditions a), b), and c) just indicated, form a hierarchy, being increasingly restrictive. At level b), a maximal solution is cited for (1) on $(a, b]$; if this is denoted $\hat{W}(t)$, then if $W(t)$ is any other solution existing on $(a, b]$, $W(t) \leq \hat{W}(t)$ everywhere. Conversely if $W \leq \hat{W}(b)$, then a solution $W(t)$ of (1) with $W(b) = W$, exists on $(a, b]$ (from Lemma 2 below).

If $\hat{Y}(t)$ is an invertible solution of $Y' = C(t)\hat{W}(t)Y$, then $\langle \hat{Y}(t), \hat{W}(t)\hat{Y}(t) \rangle$ is a solution of (2) on $(a, b]$, and is called the principal solution.

In the literature to date, principal solutions have been introduced only with $[C]$ and $[WD]$, at least, applying [Coppel [*3*], Hartman [*2*], Reid [*4*], [*14*]]. Reid's paper [*10*] is an exception, but, as is indicated below, is in a direction different from our development. A condition equivalent to our b) appears in Reid [*15*].

However we prefer to go further than that, and define a principal solution at level a), that is, given (eventually) only $[WD]$ in a neighbourhood of the point in question.

If one solution of (1) with the initial value $W(b) = W_1$, exists on $(a, b]$, so do all solutions with initial values $W(b) \leq W_1$. Furthermore if $W_1(t), W_2(t)$ exist on $(a, b]$, then so does a solution $W_3(t)$ where $W_3(t) \geq W_1(t)$, $W_3(t) \geq W_2(t)$ (this result is not proved in the text below, since it is not needed for the method of development chosen). So the set of solutions existing on $(a, b]$ (and, equivalently, their values at b) form a directed set, and even in the absence of other conditions, it is reasonable to look for a maximal element, since a directed set bounded above will indeed have a maximal element.

Such an element can be characterised in terms of infinite-valued symmetric matrices.

Infinite-valued matrices are best defined by a transformation from symmetric matrices to unitary matrices, used by Lidskii [*1*] and Atkinson [*1*], namely $L = (W+iI)(W-iI)^{-1}$ where W is any symmetric matrix. If for some vector x , $Wx = 0$, then $Lx = -x$. If W is replaced by W^{-1} then L is replaced by $-L$. Unitary matrices with unit eigenvalues are not generated by any finite symmetric matrices, but correspond (one-to-one) with the set of infinite-valued matrices which we shall introduce.

The above transformation does not preserve the ordering of matrices in a useful way. For us, infinite-valued matrices serve the same function as $\pm\infty$ in the extension of the real line, to ensure that all sets (of symmetric matrices) have upper and lower bounds. They are introduced to make the results below more easily understood; none of the proofs rely on their properties.

In fact, the extended set could be defined as the set of symmetric matrices together with the limits of all ascending and descending sequences, and the differences of the limits, or alternatively, as the set of homogeneous quadratic functions from R^n to the extended real line.

In general, the inverse of an infinite valued matrix will be finite and singular; it is this device which allows us to avoid reliance on any formal definition of infinite-valued matrices. For corresponding to each (possibly infinite-valued) maximal solution is a quite ordinary solution of the linear system (2), once again called the principal solution and denoted $\langle \hat{Y}(t), \hat{Z}(t) \rangle$. In general $\hat{Y}(t)$ is invertible iff the corresponding maximal solution is "finite-valued", that is, exists at t.

If $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ is a principal solution at a, on the interval $(a, b]$ then it has the following properties:

a) If $W(t)$ is a solution of (1) existing on $(a, b]$, then $\hat{Y}^*(t)\hat{Z}(t) - \hat{Y}^*(t)W(t)\hat{Y}(t) \geq 0$ on $(a, b]$ and $\to 0$ as $t \to a$.

b) Let $N = \hat{Z}^*(t)Y(t) - \hat{Y}^*(t)Z(t)$, where $\langle Y(t), Z(t) \rangle$ is a solution of (2), and $Y(t)$ is invertible on $(a, b]$. Then $NY^{-1}(t)\hat{Y}(t) \geq 0$ on $(a, b]$, and $\to 0$ as $t \to a$. If N is invertible, $Y^{-1}(t)\hat{Y}(t) \to 0$ as $t \to a$.

c) If $W(t)$ is any solution of (1) existing on $(a, b]$, and $\langle Y_c(t), Z_c(t) \rangle$ is the solution of (2) with $Y_c(c) = 0$,

$Z_c(c) = \hat{Z}(c) - W(c)\hat{Y}(c)$, $c \in (a, b]$ then $Y_c(t) \to \hat{Y}(t)$, $Z_c(t) \to \hat{Z}(t)$ as $c \to a$.

Our first task will be to show that, given $[R]$ on $(a, b]$, there is indeed a solution $\langle \hat{Y}, \hat{Z} \rangle$ with properties a), b), c), and that no other solution (except those obtained by post-multiplication with a constant matrix) shares them. Any one of a), b), c) would suffice to define principal solutions; we use c) and note that the property of being a principal solution at a is a local one, and once defined, the solution exists throughout I regardless of whether $[R]$ or any other conditions (even $C(t) \geq 0$) hold. When we refer to a principal solution on some interval, we really mean the solution associated with one of the endpoints; since the principal solution associated with a closed endpoint is trivial, to call a solution on a half-open interval principal will mean that it is a principal solution at the open end-point. Of course, this usage is ambiguous for open intervals, and where doubt exists, the end-point is specified.

Remarks on disconjugacy. The nature of the problem faced in this chapter may be seen more clearly by looking at the scalar version of (1) mapped onto the unit circle. Let $u(t) = \big(w(t)+i\big)\big(w(t)-i\big)^{-1}$, $v(t) = \big(z(t)+iy(t)\big)\big(z(t)-iy(t)\big)^{-1}$, where $w(t)$ is a solution of (1) and $\langle y, z \rangle$ a solution of (2). Then $u(t)$ and $v(t)$ are both solutions of

$$2u' = ia(t)(u-1)^2 - 2b(t)\big(u^2-1\big) - ic(t)(u+1)^2 , \qquad (1a)$$

and lie on the unit circle in the complex plane, as do all solutions of (1a), with initial values on the circle.

Under this transformation $y(t) = 0$ or $w(t) = \infty$ transform to $u(t) = 1$. So a solution $w(t)$ of (1) exists on an interval provided a solution $u(t)$ of (1a) exists which is not equal to 1 anywhere.

Solutions of (1a) distinct for one value of t are distinct everywhere. And $c(t) \geq 0$ ensures (by the inequalities of Chapter 2) that if $u\big(t_0\big) = 1$, then $u(t)$ proceeds in an anticlockwise direction through 1 .

Suppose a solution $u(t)$ of (1a) starts at 1 when $t = a$, travels around the circle, and equals 1 again when $t = b$. Then on $[a, b]$ every other solution must pass through 1 , and $[R]$ cannot hold. So a necessary condition for $[R]$ will forbid this behaviour on any compact subinterval, and it will be sufficient to ensure that $u(a) = u(b) = 1$ and $u(t) \neq 1$ for some t in $[a, b]$, cannot occur. For if it does then $u(t)$ must leave 1 and approach 1 in an anticlockwise fashion, and so must go around the circle.

However it is possible that $u(a) = u(b) = 1$ and $u(t) \equiv 1$ on $[a, b]$. This is not inconsistent with $[R]$ but in fact ensures that all solutions of (1) exist on $[a, b]$, regardless of initial value.

$[WD]$ is the condition forbidding the first kind of behaviour. Behaviour of the second kind is always forbidden in the literature, either by $[C]$ or a lesser condition like $M_+(a)$ (Reid [*15*]). If this is done, then one shows that $[WD]$ and $[C]$ is sufficient for $[R]$ to hold.

It is a purpose of this chapter to show that behaviour of the second kind need not be excluded in the conditions for $[R]$ to hold (Theorem 4).

Solutions of (1a) which tend to 1 at either endpoint of an interval (of any kind) are given special attention, being called principal in the domain of solutions of (2), and maximal or minimal as solutions of (1).

Basic operations on the matrix Riccati equation

Most of the work of this and the remaining chapters makes use of the fact that the Riccati equation (1) transforms into another Riccati equation under the following operations:

a) **Translation**: If $T(t)$ is a symmetric differentiable $n \times n$ matrix function, and $V(t) = W(t) + T(t)$, then

$$\begin{aligned} R[W(t)] &= A(t) + \big(V(t)-T(t)\big)B(t) + B^*(t)\big(V(t)-T(t)\big) \\ &\qquad + \big(V(t)-T(t)\big)C(t)\big(V(t)-T(t)\big) + W'(t) \\ &= R[-T(t)] + V(t)\big(B(t)-C(t)T(t)\big) + \big(B^*(t)-T(t)C(t)\big)V(t) \\ &\qquad + V(t)C(t)V(t) + V'(t) \ . \end{aligned}$$

b) **Inversion:** If $W(t)$ is invertible on an interval I , $V(t) = W^{-1}(t)$ then

$$R[W(t)] = V^{-1}(t)[-V'(t)+V(t)A(t)V(t)+B(t)V(t)+V(t)B^*(t)+C(t)]V^{-1}(t) \ .$$

So $W(t)$ is a solution of (1) iff $V(t)$ is a solution of

$$V'(t) - C(t) - B(t)V(t) - V(t)B^*(t) - V(t)A(t)V(t) = 0 \ .$$

c) **Congruence:** If $K(t)$ is an invertible solution of $K' = M(t)K$ on I , $M(t)$ some continuous matrix function, and $V(t) = K^*(t)W(t)K(t)$ then

$$\begin{aligned} &V'(t) \\ &= K^*(t)M^*(t)W(t)K(t) + K^*(t)W(t)M(t)K(t) + K^*(t)W'(t)K(t) \\ &= K^*(t)[M^*(t)W(t)+W(t)M(t)-A(t)-B^*(t)W(t)-W(t)B(t)-W(t)C(t)W(t)]K(t) \end{aligned}$$

that is,

$$\begin{aligned} &V'(t) + K^*(t)A(t)K(t) + K^*(t)[M^*(t)-B^*(t)]{K^*}^{-1}(t)V(t) \\ &\qquad + V(t)K^{-1}(t)[M(t)-B(t)]K(t) + V(t)K^{-1}(t)C(t){K^*}^{-1}(t)V(t) = 0 \ . \end{aligned}$$

The congruence transformation is used to eliminate linear terms, with $M(t) = B(t)$. Inversion is used to deal with solutions which become unbounded near a point, and translation either to eliminate constant terms (if $R[-T(t)] = 0$) or to ensure that a certain class of solutions is positive (or negative) definite on some interval. In Theorem 1, for example, all three transformations are used to reduce the Riccati equation to a very simple form with an explicit solution.

Translation and congruence preserve the ordering of solutions, and inversion inverts it, where the solutions are positive definite (or negative definite) at least.

The transformation from symmetric to unitary matrices $L = (W+iI)(W-iI)^{-1}$ mentioned earlier, is a combination of translations and an inversion. It leads to a Riccati equation (with complex coefficients) and solutions exist everywhere, since there is no boundary to the domain of definition. As mentioned earlier, ordering of solutions is not preserved in a useful form.

Finally, the operations of translation and inversion can be, and often are, carried out on solutions of (2) rather than (1). For example, if $\langle Y, Z\rangle$ is a solution of (2), and $W(t) = Z(t)Y^{-1}(t)$ exists, then

$$\big(W(t)+T(t)\big) = \big(Z(t)+T(t)Y(t)\big)Y^{-1}(t)$$

and $\big(W(t)+T(t)\big)^{-1}$, if it exists, is $Y(t)\big(Z(t)+T(t)Y(t)\big)^{-1}$. The point here is that $Y(t)$ and $Z(t)$ may both be singular, and yet an expression of the form $Y(t)\big(Z(t)+T(t)Y(t)\big)^{-1}$ may exist (see Lemma 3).

Miscellaneous lemmas

LEMMA 1. *Let A and B be two symmetric matrices with $A \geq B \geq 0$. If $B > 0$, then $B^{-1} \geq A^{-1} > 0$. Otherwise if $\eta = B\xi$ for any vector ξ, there exists a vector $\zeta : \eta = A\zeta$, and then $\eta^*(\xi-\zeta) \geq 0$.*

Proof. The second statement extends and includes the first. The proofs of both are analogous, but that of the first is easier, so it is given separately. If $A \geq B > 0$, then

$$B^{-1} - A^{-1} = B^{-1}(A-B)A^{-1} = A^{-1}(A-B)B^{-1} .$$

Substituting for B^{-1},

$$\begin{aligned} B^{-1} - A^{-1} &= \big(A^{-1}+A^{-1}(A-B)B^{-1}\big)(A-B)A^{-1} \\ &= A^{-1}(A-B)A^{-1} + A^{-1}(A-B)B^{-1}(A-B)A^{-1} \geq 0 . \end{aligned}$$

If $A \geq B \geq 0$, let N_1, N_2 be the respective null-spaces of A, B and T_1, T_2 their orthogonal complements. Then $N_2 \supseteq N_1$ and so $T_2 \subseteq T_1$.

So if $\eta = B\xi$ then $\eta \in T_2$ (since if $\varphi \in N_2$, $\varphi^*B\xi = \varphi^*\eta = 0$). Therefore $\eta \in T_1$. But T_1 is the range of A, since the range of A is included in T_1, and

$$\dim\left(T_1\right) = n - \dim\left(N_1\right) = \text{rank } A = \dim(\text{range of } A) .$$

Therefore $\exists\zeta : \eta = A\zeta$. Now

$$\begin{aligned} \eta^*(\xi-\zeta) &= \xi^*B(\xi-\zeta) = \zeta^*B(\xi-\zeta) + (\xi-\zeta)^*B(\xi-\zeta) \\ &= \zeta^*(A-B)\zeta + (\xi-\zeta)^*B(\xi-\zeta) \\ &\geq 0 . \end{aligned}$$

A bound to a solution $W(t)$ of (1) is given by

$$\psi(t) = \left(\Omega^{-1}(t)\right)^*\left[W\left(t_0\right) - \int_{t_0}^{t} \Omega^*(u)A(u)\Omega(u)du\right]\Omega^{-1}(t) \tag{6}$$

where $\Omega(t)$ is the solution of $\Omega' = B(t)\Omega$ with $\Omega\left(t_0\right) = I$. This is proved, together with a consequence, in the following lemma.

LEMMA 2. *$R[\psi(t)] \geq 0$ on I , so $\operatorname{sgn}\left(t-t_0\right)[W(t)-\psi(t)] \leq 0$ wherever $W(t)$ exists.*

If $\varphi(t)$ is a solution of $R[\varphi(t)] \leq 0$ existing in some interval $\left(a, t_0\right]$, say, and U is a symmetric matrix: $W\left(t_0\right) \leq U \leq \varphi\left(t_0\right)$ then the solution $U(t)$ of (1) with $U\left(t_0\right) = U$ exists in $\left(a, t_0\right]$, and $\psi(t) \leq U(t) \leq \varphi(t)$ on $\left(a, t_0\right]$.

Proof.

$$\begin{aligned} \psi'(t) &= -A(t) - \left(\Omega^{-1}(t)\right)^*\Omega^{*\prime}(t)\psi(t) - \psi(t)\Omega'(t)\Omega^{-1}(t) \\ &= -A(t) - B^*(t)\psi(t) - \psi(t)B(t) , \end{aligned}$$

so $R[\psi(t)] = \psi(t)C(t)\psi(t) \geq 0$ on I . And, $\psi(t)$ exists everywhere on I . The remaining statements are simple applications of Theorem 1 of Chapter 2.

LEMMA 3. *Let Y, Z be two $n \times n$ matrices for which $Z^*Y = Y^*Z$ and $Y\xi = Z\xi = 0$ only if $\xi = 0$. Then $Z - \lambda Y$ is invertible for all except at most n values of λ , and for λ large or small enough respectively, $Y^*Z - \lambda Y^*Y$ and $Y(Z-\lambda Y)^{-1}$ are non-positive or non-negative definite.*

Proof. Either $\det(Z-\lambda Y)$ is an n-th order polynomial in λ , with at most n zeros, or it is zero for all λ . We show $\det(Z-\lambda Y) \not\equiv 0$.

Let $A(\lambda) = Y^*Z - \lambda Y^*Y$, so $A(\lambda)$ is symmetric. Let N be the null-space of Y , T its orthogonal complement, S the set of unit vectors in T and Q the mapping $Q : \xi \to \xi^*Y^*Y\xi$. Then Q is continuous, S is a compact set, so $Q(S)$ is a compact set of real positive numbers and so has a least element $\varepsilon > 0$.

And there exists a number $\alpha > 0$ for which $|\eta^*Y^*Z\eta| \leq \alpha\eta^*\eta$ for any vector η .

Let ζ be any vector in R^n , with components η and ξ in N and T respectively. Then, since $Y\zeta = Y\xi$, and Y^*Z is symmetric,

$$\zeta^*Y^*Z\zeta = \xi^*Y^*Z\zeta = \xi^*Z^*Y\zeta = \xi^*Z^*Y\xi \ .$$

And $\zeta^*Y^*Y\zeta = \xi^*Y^*Y\xi$. Therefore

$$\begin{aligned} \zeta^*A(\lambda)\zeta &= \xi^*(Y^*Z-\lambda Y^*Y)\xi \\ &\leq \xi^*\xi(\alpha-\lambda\varepsilon) \\ &< 0 \quad \text{if} \quad \xi \neq 0 \quad \text{and} \quad \lambda > \frac{\alpha}{\varepsilon} \ . \end{aligned}$$

So $(Z-\lambda Y)\zeta \neq 0$ if $\xi \neq 0$ and $\lambda > \frac{\alpha}{\varepsilon}$.

If $\xi = 0$, $\zeta \in N$, so $(Z-\lambda Y)\zeta = Z\zeta \neq 0$ if $Y\zeta = 0$. Therefore

$$(Z-\lambda Y)\zeta \neq 0 \quad \text{for any} \quad \zeta \neq 0 \ , \text{ if } \quad \lambda > \frac{\alpha}{\varepsilon} \ ,$$

so $\det(Z-\lambda Y)$ cannot be identically zero for all λ .

And $A(\lambda)\zeta = 0$ if $\zeta \in N$, while if $\zeta \perp N$, $\zeta^*A(\lambda)\zeta < 0$. Therefore $\zeta^*A(\lambda)\zeta \leq 0$ for all ζ . That is, $(Y^*Z-\lambda Y^*Y)$ and

$$Y(Z-\lambda Y)^{-1} = (Z^*-\lambda Y^*)^{-1}A(\lambda)(Z-\lambda Y)^{-1}$$

are non-positive definite if $\lambda > \frac{\alpha}{\varepsilon}$. Similarly both matrices are non-negative definite if λ is sufficiently small (that is, large negative).

Existence of principal and maximal solutions

If [R] holds in an interval $(a, b]$ in I (referring to (1)), and $\langle Y(t), Z(t) \rangle$ is any solution of (2) with $Y(t)$ invertible in $(a, b]$, and $\langle Y_c(t), Z_c(t) \rangle$ is the solution of (2) with

$Y_c(c) = 0$, $Z_c(c) = Y^{*-1}(c)$, $a < c \le b$, then

DEFINITION. $\langle \hat{Y}(t), \hat{Z}(t) \rangle = \lim_{c \to a} \langle Y_c(t), Z_c(t) \rangle$ *is defined to be the principal solution of* (2) *at* a .

A similar definition gives the principal solution at a right endpoint. The fact that $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ does exist in $(a, b]$ and is a solution of (2) is established in the following theorem.

THEOREM 1. *Suppose* (1) *has a solution* $W_1(t)$ *existing on* $(a, b]$. *Then* (2) *has a principal solution* $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ *in* $(a, b]$ *and for any solution* $W(t)$ *of* (1) *existing in* $(a, b]$,

$$\hat{Y}^*(t)\hat{Z}(t) \ge \hat{Y}^*(t)W(t)\hat{Y}(t) .$$

If $\langle Y(t), Z(t) \rangle$ *is a solution of* (2) *for which* $Y(t)$ *is invertible, and if*

$$N = \hat{Z}^*(t)Y(t) - \hat{Y}^*(t)Z(t) ,$$

then

$$NY^{-1}(t)\hat{Y}(t) \to 0 \quad as \quad t \to a ,$$

If N *is invertible, then* $Y^{-1}(t)\hat{Y}(t) \to 0$ *as* $t \to a$.

Proof. From Lemma 2, if $W_2 < W_1(b)$, there exists a solution $W_2(t)$ of (1) on $(a, b]$, with $W_2(b) = W_2$.

If $W(t)$ is a solution of (1) for which $W(t) - W_2(t)$ exists and is invertible on some interval, then $V(t) = \big(W(t)-W_2(t)\big)^{-1}$ is a solution of

$$V' = C(t) + \big(B(t)+C(t)W_2(t)\big)V + V\big(W_2(t)C(t)+B(t)\big) . \qquad (7)$$

Conversely, if $V(t)$ is an invertible solution of (7) on some interval, then $V^{-1}(t) + W_2(t)$ is a solution of (1).

Let $Y_2(t)$ be an invertible solution of

$$Y_2' = \big(B(t)+C(t)W_2(t)\big)Y_2 \quad \text{on} \quad (a, b] \qquad (8)$$

and if $V(t)$ is a solution of (7), let $U(t) = Y_2^{-1}(t)V(t)Y_2^{*^{-1}}(t)$. Then

$$U'(t) = Y_2^{-1}(t)\left(C(t)+\left(B(t)+C(t)W_2(t)\right)V(t)+V(t)\left(B^*(t)+W_2(t)C(t)\right)\right)Y_2^{*^{-1}}(t)$$
$$- Y_2^{-1}(t)\left(B(t)+C(t)W_2(t)\right)V(t)Y_2^{*^{-1}}(t) -$$
$$- Y_2^{-1}(t)V(t)\left(B^*(t)+W_2(t)C(t)\right)Y_2^{*^{-1}}(t)$$
$$= Y_2^{-1}(t)C(t)Y_2^{*^{-1}}(t) \ . \tag{9}$$

Therefore

$$U(c) - U(d) = S_2(d, c)$$

where

$$S_2(d, c) = \int_d^c Y_2^{-1}(t)C(t)Y_2^{*^{-1}}(t)dt \geq 0 \ .$$

Now $V_1(t) = \left(W_1(t)-W_2(t)\right)^{-1}$ exists and is a solution of (7) on $(a, b]$, and $V_1(t) > 0$ on $(a, b]$. Therefore

$$U_1(t) = Y_2^{-1}(t)V_1(t)Y_2^{*^{-1}}(t)$$

is a solution of (9) on $(a, b]$; and

$$S_2(d, c) = U_1(c) - U_1(d) \leq U_1(c) \quad \text{if} \quad a < d \leq c \leq b \ . \tag{10}$$

Therefore, as $d \to a$, $S_2(d, c)$ is non-decreasing and bounded above. So $S_2(a, c)$ exists for each c in $(a, b]$.

Let

$$\hat{V}(t) = Y_2(t)S_2(a, t)Y_2^*(t) \ . \tag{11}$$

Then $\hat{V}(t)$ is a solution of (7), with $\hat{V}(t) \geq 0$.

Let $V(t)$ be any other solution of (7) on $(a, b]$ with $V(t) \geq 0$. Then $U(t) = Y_2^{-1}(t)V(t)Y_2^{*^{-1}}(t)$ is a solution of (9), and $U(c) = S_2(d, c) + U(d)$ if $a < c \leq d \leq b$.

So $U(c) \geq S_2(d, c)$ for all d in $(a, c]$ and so $U(c) \geq S_2(a, c)$. Therefore $V(c) \geq Y_2(c)S_2(a, c)Y_2^*(c) = \hat{V}(c)$.

Now let

$$\hat{Y}(t) = Y_2(t)\hat{U}(t) = \hat{V}(t)Y_2^{*-1}(t) \qquad (12)$$

and

$$\hat{Z}(t) = Y_2^{*-1}(t) + W_2(t)\hat{Y}(t) . \qquad (13)$$

Then

$$\begin{aligned} \hat{Y}'(t) &= Y_2'(t)\hat{U}(t) + Y_2(t)\hat{U}'(t) \\ &= B(t)\hat{Y}(t) + C(t)\left[Z_2(t)\hat{U}(t)+Y_2^{*-1}(t)\right] \\ &= B(t)\hat{Y}(t) + C(t)\hat{Z}(t) . \end{aligned}$$

And

$$\begin{aligned} \hat{Z}'(t) = &\left[Y_2^{*-1}(t)+Z_2(t)\hat{U}(t)\right]' \\ &= - \left(W_2(t)C(t)+B^*(t)\right)Y_2^{*-1}(t) + Z_2(t)Y_2^{-1}(t)C(t)\left[Y_2^{-1}(t)\right]^* \\ &\qquad\qquad - \left(B^*(t)Z_2(t)\hat{U}(t)+A(t)Y_2(t)\hat{U}(t)\right) \\ &= - B^*(t)\hat{Z}(t) - A(t)\hat{Y}(t) . \end{aligned}$$

So $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ is a solution of (2), and if $\hat{Y}(t)\xi = 0$ for some t , then $\hat{Z}(t)\xi = Y_2^{*-1}(t)\xi \neq 0$ unless $\xi = 0$. Therefore $\langle \hat{Y}, \hat{Z} \rangle$ is non-degenerate, and it is clearly isotropic; and

$$Y_2^{-1}(t)\hat{Y}(t) = \hat{U}(t) \to 0 \quad \text{as} \quad t \to a . \qquad (14)$$

Let $W_3(t)$ be some other solution of (1) existing on $(a, b]$ with $W_3(b) > W_2(b)$.

Then $W_3(t) > W_2(t)$ on $(a, b]$, and $V_3(t) = \left(W_3(t)-W_2(t)\right)^{-1}$ is a solution of (5) with $V_3(t) > 0$ on $(a, b]$. Therefore $V_3(t) \geq \hat{V}(t)$. Let $X(t) = V_3(t) - \hat{V}(t) \geq 0$. Omitting arguments for the moment,

$$\hat{V} - \hat{V}V_3^{-1}\hat{V} = XV_3^{-1}\hat{V} = \hat{V}V_3^{-1}X$$

by symmetry. Therefore

$$\begin{aligned}\hat{V} - \hat{V}V_3^{-1}\hat{V} &= \left(\hat{V}V_3^{-1}\hat{V}+XV_3^{-1}\hat{V}\right)V_3^{-1}X \\ &= \hat{V}V_3^{-1}XV_3^{-1}\hat{V} + XV_3^{-1}\hat{V}V_3^{-1}X \\ &\geq 0 .\end{aligned}$$

Therefore

$$Y_2^{-1}\left(\hat{V}-\hat{V}\left(W_3-W_2\right)\hat{V}\right)Y_2^{*-1} \geq 0 ,$$

that is,

$$Y_2^{-1}(t)\hat{Y}(t) - \hat{Y}^*(t)\left(W_3(t)-W_2(t)\right)\hat{Y}(t) \geq 0 .$$

But $Y_2^{-1}(t)\hat{Y}(t) \to 0$ as $t \to a$. Therefore

$$\hat{Y}^*(t)\left(W_3(t)-W_2(t)\right)\hat{Y}(t) \to 0 \quad \text{as} \quad t \to a$$

and

$$Y_2^{-1}(t) + \hat{Y}^*(t)W_2(t) = \hat{Z}^*(t)$$

so

$$\hat{Z}^*(t)\hat{Y}(t) - \hat{Y}^*(t)W_3(t)\hat{Y}(t) \geq 0 \quad \text{and} \quad \to 0 \quad \text{as} \quad t \to a . \qquad (15)$$

If $\langle Y_3, Z_3 \rangle$ is a solution of (2) for which $Y_3(t)$ is invertible, $Z_3(t)Y_3^{-1}(t) = W_3(t)$, then

$$\left(\hat{Z}^*(t)Y_3(t)-\hat{Y}^*(t)Z_3(t)\right)Y_3^{-1}(t)\hat{Y}(t) \geq 0 \quad \text{and} \quad \to 0 \quad \text{as} \quad t \to a ,$$

that is,

$$NY_3^{-1}(t)\hat{Y}(t) \geq 0 \quad \text{and} \quad \to 0 \quad \text{as} \quad t \to a . \qquad (16)$$

It remains to show that the solution $\langle \hat{Y}, \hat{Z} \rangle$ does not depend on the choice of solution $W_2(t)$ of (1) used to construct it. Then the requirement that $W_3(t) > W_2(t)$ in the result just established is not a restriction, since a solution $W_2(t)$ fulfilling this requirement can always be found.

Suppose two different such choices $W_1(t)$, $W_2(t)$ are made, and the resulting matrices $\hat{V}$, $\hat{Y}$ are differentiated from each other in our notation by suffices. In particular

$$Y_{i,d}(t) = Y_i(t)S_i(d,\ t)\ ,$$

$$Z_{i,d}(t) = Y_i^{*^{-1}}(t) + Z_i(t)S_i(d,\ t)\ ,\quad i = 1,\ 2\ ,\quad a < d \le b\ .$$

Then it is simple to verify that $\langle Y_{i,d},\ Z_{i,d}\rangle$ is a solution of (2). So if

$$N(d) = Y^*_{1,d}(t)Z_{2,d}(t) - Z^*_{1,d}(t)Y_{2,d}(t)$$

then $N(d)$ is independent of t and $\varepsilon\sigma = 0$, since $Y_{i,d}(d) = 0$. Therefore,

$$\hat{Y}_1^*(t)\hat{Z}_2(t) - \hat{Z}_1^*(t)\hat{Y}_2(t) = \lim_{d\to a} N(d) = 0\ .$$

Also $Y_2^*(t)\hat{Z}_1(t) - Z_2^*(t)\hat{Y}_1(t) = K$ independent of t . Therefore

$$\begin{aligned}\hat{Z}_1(t) &= Y_2^{*^{-1}}(t)Z_2^*(t)\hat{Y}_1(t) + Y_2^{*^{-1}}(t)K \\ &= W_2(t)\hat{Y}_1(t) + Y_2^{*^{-1}}(t)K\end{aligned}$$

and

$$\begin{aligned}N = 0 &= \hat{Y}_1^*(t)\left[Y_2^{*^{-1}}(t)+W_2(t)\hat{Y}_2(t)\right] - \hat{Z}_1^*(t)\hat{Y}_2(t) \\ &= \hat{Y}_1^*(t)Y_2^{*^{-1}}(t) - K^*Y_2^{-1}(t)\hat{Y}_2(t)\ .\end{aligned}$$

But $Y_2^{-1}(t)\hat{Y}_2(t) = S_2(a,\ t) = \hat{Y}_2^*(t)Y_2^{*^{-1}}(t)$ so $\hat{Y}_1(t) = \hat{Y}_2(t)K$ and

$$\begin{aligned}\hat{Z}_1(t) &= \hat{W}_2(t)\hat{Y}_1(t) + Y_2^{*^{-1}}(t)K \\ &= \left[\hat{W}_2(t)\hat{Y}_2(t)+Y_2^{*^{-1}}(t)\right]K \\ &= \hat{Z}_2(t)K\ .\end{aligned}$$

So except for post-multiplication by an arbitrary constant matrix

$\langle \hat{Y}, \hat{Z} \rangle$ is independent of the solution $W_2(t)$ chosen to construct it.

DEFINITION. *We say a matrix* $Y(t)$ *has property* $D(a, c]$ *on an interval* $(a, c]$ *of* $Y(c)\xi = i$ *only when* $Y(t)\xi = 0$ *for all* t *in* $(a, c]$. *In other words, the null space of* $Y(t)$ *is a non-expanding set.*

COROLLARY 1. *Suppose* $[R]$ *holds on* $(a, b]$, $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ *is the principal solution of* (2) *at* a , *and* $\langle Y(t), Z(t) \rangle$ *is another (non-degenerate) solution of* (2). *Let* $W_2(t)$ *be a solution of* (1) *existing on* $(a, b]$ *and* λ *be a scalar large enough that* $W_2(b) \geq -\lambda I$, $\hat{Z}(b) + \lambda\hat{Y}(b)$ *and* $Z(b) + \lambda Y(b)$ *are invertible and* $\hat{Y}(b)\{\hat{Z}(b)+\lambda\hat{Y}(b)\}^{-1} \geq 0$, $Y(b)\{Z(b)+\lambda Y(b)\}^{-1} \geq 0$. *Such* λ *exist, by Lemma* 3.

Then $Y(t)$ *has property* $D(a, c]$ *for all* c *in* $(a, b]$ *iff*

$$Y(b)\{Z(b)+\lambda Y(b)\}^{-1} \geq \hat{Y}(b)\{\hat{Z}(b)+\lambda\hat{Y}(b)\}^{-1} \geq 0 . \tag{17}$$

In this case, $Y(t)\xi = 0$ *for* $t \in (a, b]$ *implies*

$$\hat{Y}(t)\{\hat{Z}(b)+\lambda\hat{Y}(b)\}^{-1}\{Z(b)+\lambda Y(b)\}\xi = 0$$

and

$$Y(b)\xi = \hat{Y}(b)\{\hat{Z}(b)+\lambda\hat{Y}(b)\}^{-1}\{Z(b)+\lambda Y(b)\}\xi \tag{18}$$

so $\{\hat{Z}^*(b) \quad Y(b)-\hat{Y}^*(b)Z(b)\}\xi = 0$. *So* $Y(t)$ *is invertible in* $(a, b]$ *if*

$$Y(b)\{Z(b)+\lambda Y(b)\}^{-1} > \hat{Y}(b)\{\hat{Z}(b)+\lambda\hat{Y}(b)\}^{-1} \geq 0 \tag{19}$$

or if $Y^*(b)\hat{Z}(b) - Z^*(b)\hat{Y}(b)$ *is invertible, and*

$$\hat{Y}^*(b)\hat{Z}(b) - \hat{Y}^*(b)Z(b)Y^{-1}(b)\hat{Y}(b) \geq 0 . \tag{20}$$

The point of this last pair of conditions is that no particular λ is used.

Proof. Let $W_3(t)$ be the solution of (1) with $W_3(b) = -\lambda I$, and $Y_3(t)$ the solution of $Y' = \{B(t)+C(t)W_3(t)\}Y$ with $Y_3(b) = I$, and $Z_3(t) = W_3(t)Y_3(t)$. Then $W_3(t)$ exists on $(a, b]$

by Lemma 2, and $\langle Y_3(t), Z_3(t)\rangle$ is a solution of (2).

Let

$$N = Y_3^*(t)Z(t) - Z_3^*(t)Y(t) = Z(b) + \lambda Y(b)$$

and

$$\hat{N} = Y_3^*(t)\hat{Z}(t) - Z_3^*(t)\hat{Y}(t) = \hat{Z}(b) + \lambda\hat{Y}(b) .$$

Let

$$U(t) = Y_3^{-1}(t)Y(t)N^{-1} , \quad \hat{U}(t) = Y_3^{-1}(t)\hat{Y}(t)\hat{N}^{-1} .$$

Then

$$\begin{aligned} U'(t) &= -Y_3^{-1}(t)\{B(t)+C(t)W_3(t)\}Y(t)N^{-1} + Y_3^{-1}(t)\{B(t)Y(t)+C(t)Z(t)\}N^{-1} \\ &= Y_3^{-1}(t)C(t)Y_3^{*-1}(t) . \end{aligned}$$

Similarly $\hat{U}'(t) = Y_3^{-1}(t)C(t)Y_3^{*-1}(t)$; and

$$U(b) = Y(b)\{Z(b)+\lambda Y(b)\}^{-1} \geq 0 ,$$

and $\hat{U}(b) \geq 0$ also. From Theorem 1, $\hat{U}(t) \geq 0$, and $\hat{U}(t) \to 0$ as $t \to a$. So if $U(b) \geq \hat{U}(b)$, then $U(t) = U(b) + \hat{U}(t) - \hat{U}(b) \geq 0$ in $(a, b]$. And if $U(c)\xi = 0$ then $U(t)\xi = 0$ for all $t \leq c$, since $U(t)$ is a non-decreasing non-negative function.

Conversely, suppose that $U(t)$ is not non-negative in $(a, b]$. Then there is a maximum compact interval $[c, b]$ on which it is non-negative. At c , there are say, r positive eigenvalues of $U(t)$, and this is true in any small enough neighbourhood of c . So if negative eigenvalues of $U(t)$ appear in every left neighbourhood of c , the nullspace of $U(t)$, and hence of $Y(t)$, must diminish. So (17) is established, since $Y(t)$ has property $D(a, c]$ iff $U(t)$ has it.

If $\hat{U}(t)\xi = 0$, and $U(b) \geq \hat{U}(b)$, then $U(t)\xi = 0$, so $\{U(b)-\hat{U}(b)\}\xi = \{U(t)-\hat{U}(t)\}\xi = 0$; that is, if $\xi = N\eta$, then

$$\begin{aligned} Y(t)\eta = 0 \quad \text{iff} \quad & \{U(b)-\hat{U}(b)\}\xi \\ & = \{Y(b)N^{-1}-\hat{Y}(b)\hat{N}^{-1}\}N\eta = 0 \end{aligned}$$

proving (18). So if $Y(b) - \hat{Y}(b)\hat{N}^{-1}N$ is invertible and non-negative $Y(t)$ must be invertible everywhere. And

$$\begin{aligned} Y(b) - \hat{Y}(b)\hat{N}^{-1}N &= \hat{N}^{*-1}\left(\hat{N}^*Y(b)-\hat{Y}^*(b)N\right) \quad \text{since } \hat{Y}(b)\hat{N}^{-1} \text{ is symmetric} \\ &= \hat{N}^{*-1}\left(\hat{Z}^*(b)Y(b)-\hat{Y}^*(b)Z(b)\right) . \end{aligned}$$

So $Y(t)$ is invertible if $\hat{Z}^*(b)Y(b) - \hat{Y}^*(b)Z(b)$ is invertible and $U(b) \geq \hat{U}(b)$.

But, using the same kind of procedure as in the proof of Lemma 1,

$$\begin{aligned} U(b) - \hat{U}(b) = {} & (\hat{N}^*)^{-1}\left(\hat{Y}^*(b)N-\hat{N}^*Y(b)\right)Y^{-1}(b)\hat{Y}(b)\hat{N}^{-1} \\ & + (N^*)^{-1}\left(\hat{Y}^*(b)N-\hat{N}^*Y(b)\right)Y^{-1}(b)(\hat{N}^*)^{-1}\left(N^*\hat{Y}(b)-Y^*(b)\hat{N}\right)\hat{N}^{-1} \\ \geq {} & 0 \end{aligned}$$

if both $N^*Y(b) \geq 0$, which is true if λ is large, and if

$$\left(\hat{Y}^*(b)N-\hat{N}^*Y(b)\right)Y(b)^{-1}\hat{Y}(b) = \hat{Y}^*(b)\hat{Z}(b) - \hat{Y}^*(b)Z(b)Y^{-1}(b)\hat{Y}(b) \geq 0$$

proving (20). (19) is the statement that $U(b) > \hat{U}(b)$.

Digression on extended (infinite-valued) matrices

The result of Corollary 1 is most naturally expressed in terms of extended (infinite-valued) matrices, with the aid of which the existence of a maximal solution can be re-asserted. If A is a symmetric linear operator mapping its domain T in R^n into itself, it is said to represent an extended matrix on R^n , denoted $\{A\}$. The images under $\{A\}$ of points not in T are not defined, but could be assumed infinite, hence "infinite-valued".

For two such extended matrices we say $\{A\} \geq \{B\}$ if the domain of $\{B\}$ includes the domain of $\{A\}$ and for all vectors ξ in the domain of A , $(\xi, B\xi) \geq (\xi, A\xi)$, or, alternatively, if the domain of A includes the domain of B and on the domain of B the same inequality holds.

In particular, any symmetric matrix has an extended matrix inverse. For if S is a symmetric matrix with null-space N and if T is the orthogonal complement of N , then S maps T bijectively and linearly onto itself, and N to zero. So $\{S^{-1}\}$ is the inverse mapping from T to T .

Lemma 1 says that if P and Q are symmetric matrices, and $P \geq Q \geq 0$, then $\{Q^{-1}\} \geq \{P^{-1}\} \geq 0$.

So in the result of Corollary 1, let $\hat{V} = \hat{Y}(b)\left(\hat{Z}(b)+\lambda\hat{Y}(b)\right)^{-1}$, $V = Y(b)\left(Z(b)+\lambda Y(b)\right)^{-1}$.

Let $\hat{N}$, N be the null-spaces of $\hat{V}$ and V respectively, and $\hat{T}$ and T the complements of $\hat{N}$ and N.

Then if $\eta \in \hat{T}$, $\eta \neq 0$, there is a unique vector ξ in $\hat{T} : \eta = \hat{V}\xi$. Then $\left(\hat{Z}^*(b)+\lambda\hat{Y}^*(b)\right)\eta = \hat{Y}^*(b)\xi$ since $\hat{V}$ is symmetric, or $\hat{Z}^*(b)\eta = \hat{Y}^*(b)(\xi-\lambda\eta)$.

By Lemma 1, there exists a vector $\zeta \in T : \eta = V\zeta$ and ζ is unique in T. And similarly $Z^*(b)\eta = Y^*(b)(\zeta-\lambda\eta)$. And again by Lemma 1, $\eta^*(\xi-\zeta) \geq 0$.

Let $\hat{W}$ be the mapping $\hat{W} : \eta \to \xi - \lambda\eta$. Then if $\eta \in \hat{T}$, $(\eta, \hat{W}\eta) = (\eta, \xi) - \lambda(\eta, \eta) = (\hat{W}\eta, \eta)$ so $\hat{W}$ is a symmetric linear mapping of $\hat{T}$ onto itself; its domain is $\hat{T}$.

Let W be the mapping $W : \eta \to \zeta - \lambda\eta$, mapping $\hat{T}$ onto a subspace of T. Then for all η in $\hat{T}$, $\eta^*\left((\xi-\lambda\eta)-(\zeta-\lambda\eta)\right) \geq 0$ that is, $(\eta, \hat{W}\eta) - (\eta, W\eta) \geq 0$. Therefore $\{\hat{W}\} \geq \{W\}$ and $\{\hat{W}\}$ is maximal.

A real symmetric matrix B is a special case of an extended matrix, and if $\{A\}$ is an extended matrix with domain T, and on T $(\xi, A\xi) > \xi^*B\xi$, then it can be said that $\{A\} > B$.

If in Corollary 1, we note that $\hat{T}$, the domain of $\{\hat{W}(b)\}$, is the range of $\hat{V}$, which is the range of $\hat{Y}(b)$, then the sufficient condition that a solution $W(t)$ of (1) exists on $(a, b]$ is that $\{\hat{W}(b)\} > W(b)$. So the set of extended matrices $\{W\}$ for which $\{\hat{W}(b)\} \geq \{W\}$ is in a sense, the closure of the translated cone of initial values of solutions of (1) existing on $(a, b]$.

Reid [10] has used a different approach to the problem of determining the principal solutions of (2) in the absence of the controllability condition [C]. In effect, in our approach whenever the inverse of a singular matrix is required, we first translate it, and deal with a different but still useful image after inversion.

Lemma 3 ensures that an appropriate translation can always be found.

Reid, in a somewhat similar situation, used Moore generalised inverses. In our extended matrix approach we make no use of the inverse of a singular matrix on its null-space; the Moore inverse, however, maps the null-space onto zero. This procedure clearly introduces a major discontinuity; for example the inverse of the scalar 0 is 0, and $\left(0^{-1}+1\right)^{-1} = 1$. If in our approach we had sought a value for $\left(0^{-1}+1\right)^{-1}$ we would have let $Y = 0$, $Z = I$, so $0 = YZ^{-1}$, and $\left(0^{-1}+1\right)^{-1} = Y(Y+Z)^{-1} = 0$.

There is a real difference in the results; for example, suppose in (2), $A(t) \equiv B(t) \equiv C(t) \equiv 0$ on $(0, 1]$. According to Reid's results, any solution $\langle K, WK \rangle$ where W is symmetric and K invertible, is a principal solution, and no others. In our approach, any solution $\langle 0, K \rangle$, with K invertible, is a principal solution, and no others. In this example the two sets of principal solutions are disjoint, and Reid's set is a much larger one, including almost all solutions. Ours is unique, except for post-multiplication by an invertible $n \times n$ matrix, and has the properties indicated in Theorem 1. It corresponds to the maximal "solution" $W = \infty$, of (1).

COROLLARY 2. *If* $[R]$ *holds in* $(a, b]$, *and* $\langle Y_a, Z_a \rangle$ *is the principal solution at* a, *then the null space of* $Y_a(t)$ *is non-expanding; that is,* $D(a, c]$ *holds if* $a < c \le b$.

Proof. This is a special case of (17) above.

COROLLARY 3. *If* $[a, b] \subseteq I$, *then the principal solution* $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ *of* (2) *at* a *exists, and* $\hat{Y}(a) = 0$, $\hat{Z}(a)$ *is invertible. So* $\hat{U}(a) = \hat{V}(a) = 0$.

Proof. $[R]$ must hold in some non-empty interval $[a, c]$, so a principal solution does exist.

Let $W(t)$ be any solution of (1) with $W(a) = W$. Then by Theorem 1,

$$\lim_{t \to a} \left\{\hat{Y}^*(t)\hat{Z}(t)-\hat{Y}^*(t)W(t)\hat{Y}(t)\right\} = \hat{Y}^*(a)\hat{Z}(a) - \hat{Y}^*(a)W\hat{Y}(a) \ge 0$$

for all W. Therefore $\hat{Y}(a) = 0$.

If $\hat{Z}(a)$ is not invertible, the solution $\langle \hat{Y}, \hat{Z} \rangle$ is degenerate.

COROLLARY 4. *[R] ⇒ [WD] on any interval.*

Proof. Suppose $[a, b] \subseteq J$, J any interval on which $[R]$ holds, and $\langle Y_a, Z_a \rangle$ is the solution of (2) with $Y_a(a) = 0$, $Z_a(a) = I$. Then $\langle Y_a, Z_a \rangle$ is a principal solution at a , and by Corollary 2, $D(a, b]$ holds, so $Y_a(b)\xi = 0$ only if $Y_a(t)\xi = 0$ on $(a, b]$. So a solution of (2a) has $y(a) = y(b) = 0$ only if $y(t) = 0$ on $[a, b]$, so $[WD]$ holds.

COROLLARY 5. *If $[R]$ holds on $(a, b]$, then $\hat{Y}(b)$ is invertible iff $P(a, b]$ holds. Consequently $\hat{Y}(t)$ is invertible everywhere on $(a, b]$ iff $M_+(a)$ holds. Then $\hat{W}(t) = \hat{Z}(t)\hat{Y}^{-1}(t)$ is a maximal solution of* (1) *on* $(a, b]$.

Proof. By Corollary 2, if $\hat{Y}(b)\xi = 0$, then $y(t) = \hat{Y}(t)\xi = 0$ on $(a, b]$, $z(t) = \hat{Z}(t)\xi$ is a solution of

$$z' = -B^*(t)z - A(t)y(t) = -B^*(t)z ,$$

so $P(a, b]$ fails.

Conversely, if $P(a, b]$ fails to hold, then there is a non-trivial solution $\langle 0, z(t) \rangle$ of (2a). Let $\langle Y_1(t), Z_1(t) \rangle$ be a solution of (2) with $Y_1(t)$ invertible on $(a, b]$ and $N = Y_1^*(b)\hat{Z}(b) - Z_1^*(b)\hat{Y}(b)$ invertible. Then

$$Y_1^*(t)z(t) = Z_1^*(t)y(t) + \xi_1 = \xi_1 ,$$

where ξ_1 is constant. And $\hat{Y}^*(t)z(t) = \hat{\xi}$, another constant. So $\hat{\xi} = \hat{Y}^*(t)\left(Y_1^*(t)\right)^{-1}\xi_1$ and the right-hand side tends to zero, as $t \to a$ by Theorem 1. So $\hat{\xi} = 0$. Therefore $\hat{Y}^*(t)z(t) = 0$ and $\hat{Y}(t)$ is singular everywhere.

That $M_+(a) \Leftrightarrow \hat{Y}(t)$ invertible on $(a, b]$ is an obvious consequence. Then if $W(t)$ is any solution existing on $(a, b]$, by Theorem 1,

$$\hat{Y}^*(t)\hat{W}(t)\hat{Y}(t) \geq \hat{Y}^*(t)W(t)\hat{Y}(t) \quad \text{on} \quad (a, b]$$

or $\hat{W}(t) \geq W(t)$.

Remark. The usual definition of a principal solution $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ (Hartman [2], Reid [4], Coppel [1]) requires that $\hat{Y}(t)$ be invertible and that

$$\left(\int_c^b \hat{Y}^{-1}(t) C(t) \hat{Y}^{*-1}(t) dt \right)^{-1} \to 0$$

as $c \to a$. We have naturally used another definition because $\hat{Y}(t)$ in our approach need not be invertible. But it is of interest to see what has become of this property.

Let $W_1(t)$ be a solution of (1) existing on $(a, b]$, $Y_1(t)$ an invertible solution of $Y' = \big(B(t)+C(t)W_1(t)\big)Y$, and $\langle Y_2(t), Z_2(t) \rangle$ a solution of (2) for which

$$N = Z_2^*(t)Y_1(t) - Y_2^*(t)W_1(t)Y_1(t)$$

is invertible. Let

$$P(t) = NY_1^{-1}(t)Y_2(t) = Z_2^*(t)Y_2(t) - Y_2^*(t)W_1(t)Y_2(t) .$$

Then

$$\begin{aligned} P'(t) &= -NY_1^{-1}(t)\big(B(t)+C(t)W_1(t)\big)Y_2(t) + NY_1^{-1}(t)\big(B(t)Y_2(t)+C(t)Z_2(t)\big) \\ &= NY_1^{-1}(t)C(t)Y_1^{*-1}(t)\big(-Y_1^*(t)W_1(t)Y_2(t)+Y_1^*(t)Z_2(t)\big) \\ &= NY_1^{-1}(t)C(t)Y_1^{*-1}(t)N^* . \end{aligned}$$

If in addition $Y_2(t)$ is invertible

$$P'(t) = NY_1^{-1}(t)Y_2(t)Y_2^{-1}(t)C(t)Y_2^{*-1}(t)Y_2^*(t)Y_1^{*-1}(t)N^* ,$$

$P(t)$ is symmetric, so

$$P'(t) = P(t)Y_2^{-1}(t)C(t)Y_2^{*-1}(t)P(t)$$

or

$$\big(P^{-1}(t)\big)' = -Y_2^{-1}(t)C(t)Y_2^{*-1}(t) .$$

Let

$$S_i(t) = \int_t^b Y_i^{-1}(u)C(u)Y_i^{-1*}(u)du \ , \quad i = 1, 2 \ .$$

Then

$$P(b) - P(t) = NS_1(t)N^*$$

and

$$P^{-1}(t) = P^{-1}(b) + S_2(t) \ .$$

Now suppose $\langle Y_2, Z_2 \rangle$ is a principal solution, with $Y_2(b)$ invertible. Then $S_2(t)$ exists near b , and so do $P^{-1}(t), P^{-1}(b)$, and $P(b) > 0, P(t) \geq 0$.

Then if $Y_2(t)$ is invertible for all t in $(a, b]$, from Theorem 1, $P(t) \to 0$ as $t \to a$, so $S_2(t) \to \infty$. If $Y_2(t)$ is not always invertible, we can still define $S_2(t)$ as an extended matrix, and it is still then true that $\{S_2(t)\} \to \infty$, that is, for any scalar α , there exists $c : \{S_2(t)\} \geq \alpha I$ on $(a, c]$.

Remark. $M_+(a)$ is the strongest condition that we need for desirable properties of solutions of (1) on a single interval. If however we want to consider classes of intervals, conditions $[C]$ and $[D]$ can be useful. Also they have been used elsewhere, so we shall digress a little to consider their implications here.

a) $[D] \Rightarrow [C]$.

If $z(t)$ is a function such that $z'(t) = B(t)z(t)$, $C(t)z(t) \equiv 0$ on $[a, b]$, then $\langle 0, z(t) \rangle$ is a solution of

$$y' = B(t)y + C(t)z \ , \quad z' = -A(t)y - B^*(t)z \ ,$$

with $y(a) = y(b) = 0$. Consequently if $[D]$ holds, $z(t) \equiv 0$, so $[C]$ holds.

b) $[C]$ *and* $[WD] \Rightarrow [D]$.

If $\langle y(t), z(t) \rangle$ is a solution of (20) with $y(a) = y(b) = 0$, and $[C]$ and $[WD]$ hold, then $y(t) \equiv 0$ on $[a, b]$. Then

$$z'(t) = -B^*(t)z(t) \ , \quad C(t)z(t) = y'(t)-B(t)y(t) \equiv 0 \ .$$

Therefore $z(t) \equiv 0$ on $[a, b]$, so $[D]$ holds.

c) *$[D]$ on $(a, b]$ implies $[R]$ on (a, b) .*

Proof. Let $\langle Y(t), Z(t)\rangle$ be the solution of (2) with $Y(b) = 0$, $Z(b) = I$. Then $Y(t)$ is invertible in (a, b) , otherwise if $Y(t)\xi = 0$, then $Y(u)\xi = Z(u)\xi = 0$ in $[t, b]$, so $\xi = 0$.

So $W(t) = Z(t)Y^{-1}(t)$ exists in (a, b) .

Theorem 4 below strengthens this result.

d) *On an open interval $[D] \Rightarrow [R]$.*

Proof. If $[D]$ holds on an open interval (a, b) it holds on all subintervals $(a, c]$. From the previous result c), $[R]$ holds on (a, c) , and a principal solution at a , $\langle \hat{Y}(t), \hat{Z}(t)\rangle$ exists on (a, c) . And $[D] \Rightarrow [C] \Rightarrow M_+(a)$ so $\hat{Y}(t)$ is invertible on (a, c) , by Corollary 5.

Suppose d is the least point in (a, b) for which $\hat{Y}(d)$ is singular.

Then, from Corollary 2, $[R]$ does not hold in $(a, d]$. But $[R]$ holds in (a, c) if $a < c < b$. So there is no such point d , and so $\hat{Y}(t)$ is invertible everywhere in (a, b) . So $\hat{Z}(t)\hat{Y}^{-1}(t)$ exists and is a solution of (1) everywhere in (a, b) , so $[R]$ holds.

e) *On a compact interval $[D] \Rightarrow [R]$.*

This is the last statement about $[D]$ that was foreshadowed earlier. It is a special case of the next theorem, which is important because it relates weak disconjugacy directly with $[R]$, and hence with the existence of a principal solution. The following lemma is necessary for its proof.

Existence of Invertible Solutions inferred from Disconjugacy

LEMMA 4. *Let $J = (a, c]$, or $[a, c]$. If there is a non-degenerate solution $\langle Y, Z\rangle$ of (2), where $Y(t)$ has a non-expanding null-space on J , (so that $Y(t)\xi = 0$ only if $Y(u)\xi = 0$ for $u \le t$) then $[R]$ holds on J .*

Proof. Let

$$\Omega(t) \text{ be a fundamental matrix of } y' = B(t)y \ , \tag{21}$$

and

$$\varphi(t) = -\int_t^c \Omega^*(u)A(u)\Omega(u)du \ , \tag{22}$$

$$D(t) = \Omega^{-1}(t)C(t)\Omega^{*-1}(t) \ , \tag{23}$$

$$Y_1(t) = \Omega^{-1}(t)Y(t) \tag{24}$$

$$Z_1(t) = \Omega^*(t)Z(t) + \varphi(t)Y_1(t) \ . \tag{25}$$

Then $Z_1^*(t)Y_1(t)$ is symmetric and if $Y_1(t)\xi = Z_1(t)\xi = 0$ then $Y(t)\xi = Z(t)\xi = 0$, so $\xi = 0$, for any t in J , and so for $t = c$. Therefore by Lemma 3, there is a number λ for which $Z_1(c) + \lambda Y_1(c)$ is invertible and $Y_1(c)\left(Z_1(c)+\lambda Y_1(c)\right)^{-1} \geq 0$.

Let

$$Y_c(t) = Y_1(t) = \Omega^{-1}(t)Y(t) \ ,$$

$$Z_c(t) = Z_1(t) + \lambda Y_1(t) = \Omega^*(t)Z(t) + \left(\varphi(t)+\lambda I\right)\Omega^{-1}(t)Y(t) \ ,$$

$$\theta(t) = \varphi(t) + \lambda I \ .$$

Then $\langle Y_c(t), Z_c(t)\rangle$ is a non-degenerate solution of the Hamiltonian system

$$Y_c'(t) = -D(t)\theta(t)Y_c(t) + D(t)Z_c(t) \ , \tag{26}$$

$$Z_c'(t) = \theta^*(t)Y_c'(t) \ . \tag{27}$$

The nullspace of $Y_c(t)$ is the same as that of $Y(t)$; we denote it $N(t)$ and its dimension $d(t)$. Then $d(t)$ is a non-increasing integer-valued function, which can change at no more than n points $a < c_h < c_{h-1} < \ldots < c_1 < c$, $h \leq n$. Let $T(t)$ be the orthogonal complement of $N(t)$ at each point.

Then $Z_c^*(c)Y_c(c) \geq 0$, and its nullspace is $N(c)$. Let

$$\lambda(t) = \min_{\substack{\xi\in T(t) \\ |\xi|=1}} \xi^* Z_c^*(t) Y_c(t)\xi \ .$$

Then $\lambda(c) > 0$, and $\lambda(t)$ is continuous on $(c_1, c]$, since $T(t)$ does not change. Let d be the greatest point in $(c_1, c]$ for which $\lambda(d) = 0$, so $\lambda(t) > 0$ in $(d, c]$. Then there is some $\xi \in T(c)$, $|\xi| = 1$, for which, $Z_c^*(d)Y_c(d)\xi = 0$.

If $Z_c(d)$ is not invertible, there is a point b in $[d, c)$ for which $Z_c(b)$ is singular, and $Z_c(t)$ is nonsingular in $(b, c]$.

Then $V(t) = Y_c(t)Z_c^{-1}(t)$ is a solution of

$$V' = (V\theta^*(t)-I)D(t)(\theta(t)V-I) \geq 0 \tag{28}$$

and $0 \leq V(t) \leq V(c)$ on $(b, c]$, since $V(c) \geq 0$ and the nullspace of $V(t)$ does not change on $(b, c]$. So $V(t)$ can be continued to b .

And $Z_c(t)$ is a solution of $Z_c' = -\theta^*(t)[D(t)+D(t)\theta(t)V(t)]Z_c$ on $[b, c]$, with $Z_c(c)$ invertible, so $Z_c(b)$ is invertible. So no such b exists, and $Z_c(d)$ is invertible. But $Z_c^*(d)Y_c(d)\xi = 0$, so $Y_c(d)\xi = 0$. Therefore, $\xi \in N(d) = N(c)$. But by assumption $\xi \in T(c)$ and $\xi \neq 0$. This is a contradiction, so there cannot be any such point d , and so $\lambda(t) > 0$ on $(c_1, c]$. Therefore $Y_c^*(t)Z_c(t) \geq 0$ on $(c_1, c]$ so $Y_c^*(c_1)Z_c(c_1) \geq 0$. The above argument can be repeated, at most n times, to show that on $(c_2, c_1]$, $(c_3, c_2]$ etc, $Z(t)$ is invertible. Therefore, $V_c(t) = Y_c(t)Z_c^{-1}(t) \geq 0$ exists and is a non-negative solution of (28).

Let $V(t)$ be a solution of (28) with $V(c) > V_c(c)$. Then $V(t)$ exists on J , and $V(c) \geq V(t) > V_c(t) \geq 0$ on J . So $W(t) = \Omega^{*-1}(t)\left[V^{-1}(t)-\theta(t)\right]\Omega^{-1}(t)$ exists and is differentiable on J , and can be verified to be a solution of (1).

THEOREM 2. *On a compact interval* $[a, b]$, $[WD] \Rightarrow [R]$.

Proof. Let $\langle Y_0, Z_0 \rangle$ be the solution of (2) with $Y_0(a) = 0$, $Z_0(a) = I$. Then $Y_0(t)$ has a non-expanding nullspace on $[a, b]$, if $[WD]$ applies. By Lemma 4, $[R]$ holds on $[a, b]$. QED

In Reid [*14*] an equivalent result is obtained by variational means (for example, Chapter VII, Theorem 5.1).

THEOREM 3. *If* $[WD]$ *holds on* $(a, b]$ *then for some* c *in* $(a, b]$, $[R]$ *holds in* $(a, c]$ *; and* $[WD]$ *on* $[b, a) \Rightarrow [R]$ *on* $[c, a)$ *for some* c *in* $[b, a)$.

Proof. Let $\langle Y_0, Z_0 \rangle$ be the solution of (2) with $Y_0(b) = 0$, $Z_0(b) = I$. Then the nullspace of $Y_0(t)$ is non-contracting on $(a, b]$, and so can only change finitely often, not more than n times. Therefore there is an interval $(a, c]$ on which it is constant, with its minimum dimension. Being constant, it is non-expanding also, so Lemma 4 applies, and $[R]$ holds on $(a, c]$. QED

The second statement in the theorem can be deduced by a symmetry argument.

Theorem 3 indicates that a principal solution $\langle \hat{Y}, \hat{Z} \rangle$ at a can be defined in an interval $(a, b]$ say, provided only that $[WD]$ holds in some right neighbourhood of a . This observation allows an extension of Theorem 3 (see Theorem 4).

DEFINITION. *A right* conjugate point *of a point* a *is defined as the first point* $b > a$ *at which the nullspace of the principal solution* $\langle \hat{Y}_a, \hat{Z}_a \rangle$ *at* a *expands. A left conjugate point is defined equivalently.*

This definition, which differs from the usual one, applies also to end-points of open and half-open intervals, provided $[WD]$ holds in an appropriate neighbourhood of the endpoint in question.

DEFINITION. Extended weak disconjugacy $[EWD]$ *holds for* (2) *on an interval* J *if there are two (non-degenerate) solutions* $\langle Y_1, Z_1 \rangle$ *and* $\langle Y_2, Z_2 \rangle$ *of* (2) *for which the nullspaces of* $Y_1(t)$ *and* $Y_2(t)$ *are non-expanding and non-contracting sets respectively as* t *increases on* J .

This form of definition is adopted because it makes no use of concepts resulting from earlier theorems. However, the following equivalent version is more informative.

LEMMA 5. *[EWD] holds on J iff [WD] holds and any open endpoints of J do not have conjugate points in J.*

Proof. If the second statement holds, then the principal solutions associated with each endpoint exist by Theorem 3, and satisfy the requirements of the definition of [*EWD*] .

Conversely, if [*EWD*] holds, then [*R*] holds on any compact or half-open subinterval of J . So [*WD*] holds on any compact subinterval, and therefore, from its definition, throughout J .

If J is open at its left endpoint a , then for any c in J , [*R*] holds on $(a, c]$, so the nullspace of $Y_a(t)$, where $\langle Y_a, Z_a \rangle$ is the principal solution of (2) at a , is non-expanding. So no point c in J is conjugate to a . A similar argument applies at the right end-point, if it is open.

Remark. The definition of conjugate point above, strictly speaking, generalises only the concept of nearest conjugate point. This is enough for our needs. The extension of [*WD*] to [*EWD*] makes a real difference even in familiar cases. For example, the system $y' = z$, $z' = -y$ obeys [*WD*] (that is, is disconjugate) on $(0, \pi)$ or $(0, \pi]$ but not on $[0, \pi]$. It obeys [*EWD*] (and [*R*]) on $(0, \pi)$ but not on $(0, \pi]$ or $[0, \pi]$, since π is a conjugate point of 0 .

A conjugate point of an open end-point may not just be the limit of the conjugate points of points in the interval. For example, if the dimension $n = 1$, $B(t) = 0$ everywhere, $A(t) = C(t) = 0$ if $t \geq 0$, $A(t) = C(t) = \dfrac{2}{1+t^2}$ if $t < 0$ (the discontinuity at 0 does not matter), then no point in $(-\infty, \infty)$ has a conjugate point. But 0 is a conjugate point of $-\infty$. [*WD*] holds on $(-\infty, \infty)$, but [*EWD*] and [*R*] fail on $(-\infty, 0]$.

THEOREM 4. *On any interval J , [R] $\Longleftrightarrow$ [EWD] .*

Proof. If $[R]$ holds on J, then by Theorem 1 and its Corollary 2, the two principal solutions at the left and right end points exist and have, respectively, non-increasing and non-decreasing null-spaces on J. So $[EWD]$ holds on J.

Conversely, for compact and half-open intervals, $[EWD] \Rightarrow [R]$ from Lemma 4.

The case of an open interval (a, b) is more difficult. Let $\langle Y_a(t), Z_a(t)\rangle$, $\langle Y_b(t), Z_b(t)\rangle$ be the principal solutions of (2) at a and b respectively. Then if $[EWD]$ holds, the nullspaces of $Y_a(t)$, $Y_b(t)$ are respectively non-expanding and non-contracting as t increases. Since each nullspace can only change n times, there is a point c for which no changes in either nullspace occur in $(a, c]$.

We abbreviate Y_a, Z_a, for $Y_a(c)$, $Z_a(c)$ etc. By Lemmas 2 and 3, and knowing that $[R]$ holds on $(a, c]$ and $[c, b)$, a number λ exists large enough so that all the following are true:

> $Z_a + \lambda Y_a$, $Z_a - \lambda Y_a$, $Z_b + \lambda Y_b$, $Z_b - \lambda Y_b$ are all invertible;
>
> $Y_a\left(Z_a+\lambda Y_a\right)^{-1} \geq 0$, $Y_a\left(Z_a-\lambda Y_a\right)^{-1} \leq 0$, $Y_b\left(Z_b+\lambda Y_b\right)^{-1} \geq 0$, $Y_b\left(Z_b-\lambda Y_b\right)^{-1} \leq 0$, and the solutions $W_1(t)$, $W_2(t)$ of (1) with $W_1(c) = \lambda I$, $W_2(c) = -\lambda I$ exist on $[c, b)$, $(a, c]$ respectively.

Let $Y_i(t)$, $i = 1, 2$ be the solutions of $Y' = C(t)W_i(t)Y$ with $Y_i(c) = I$. Then $Y_i(t)$ exists and is invertible on $[c, b)$ for $i = 1$ and on $(a, c]$ if $i = 2$, and if $Z_i(t) = W_i(t)Y_i(t)$ then $\langle Y_i(t), Z_i(t)\rangle$ are solutions of (2) for $i = 1, 2$.

Let $U_b = Y_b\left(Z_b+\lambda Y_b\right)^{-1} \geq 0$, $U_a = Y_a\left(Z_a+\lambda Y_a\right)^{-1} \geq 0$. Then from Corollary 1 to Theorem 1, since $Y_b(t)$ obeys $D(a, t]$ for t in $(a, c]$, $U_b \geq U_a \geq 0$.

Let x_i , $i = 1 \ldots h$ be an orthonormal basis for the nullspace of U_b , and let $Q = \sum_{i=1}^{h} x_i x_i^*$. Then $U_b Q = U_a Q = 0$, and $Q^2 = Q$. If $U_b x = 0$, then $Qx = x$.

Let $S = \frac{1}{2}(U_a + U_b)$ and $U = S + rQ$ for some positive constant r . Let $\langle Y, Z \rangle$ be the solution of (2) with $Y(c) = U$, $Z(c) = I - \lambda U$. Then $U > 0$, so $Y(c)$ is invertible. We shall show that $Y(t)$ is invertible on (a, b) , proving [R] .

Suppose for d in $[c, b)$, $Y(d)\xi = 0$, and $\xi \neq 0$. But

$$Y_b^* Z_b - Y_b^* Z(c) Y^{-1}(c) Y_b = U_b - U_b U^{-1} U_b$$

and $0 < U \leq U_b + Q$, and $(U_b + Q)^{-1} U_b = I - Q$. So

$$\begin{aligned} U_b - U_b U^{-1} U_b &\leq U_b - U_b (U_b + Q)^{-1} U_b \quad \text{from Lemma 1} \\ &\leq 0 \end{aligned}$$

since $U_b Q = 0$. Therefore,

$$Y_b^* Z_b - Y_b^* Z(c) Y^{-1}(c) Y_b \leq 0 \ ,$$

so, applying a dual version of (18) in Corollary 1,

$$\big(Y_b^* Z(c) - Z_b^* Y(c)\big)\xi = 0$$

that is, $(U_b - U)\xi = 0$. Then from the definition of U , $U_b \xi = U_a \xi + 2rQ\xi$. Multiplying by Q shows $Q\xi = 0$, so $U_b \xi = U_a \xi$. Then

$$Y_a(c)\xi = U_a \xi = Y_b(c)\xi \ ,$$

$$Z_a(c)\xi = (I - \lambda U_a)\xi = Z_b(c)\xi \ .$$

So $\langle Y_a(t)\xi, Z_a(t)\xi \rangle$ is the same solution of (2a) as $\langle Y_b(t)\xi, Z_b(t)\xi \rangle$ and so $Y_a(d)\xi = 0$. But $Y_a(t)$ has a non-expanding nullspace, so $Y_a(c)\xi = U_a \xi = 0$. Therefore, $U\xi = 0$. But U is invertible. So $Y(t)$ is invertible on $[c, b)$. The proof for $(a, c]$ is similar, and a little easier.

Tests for the existence and absence of disconjugacy

Condition $[R]$ for the Riccati equation (1) is a sufficient condition for disconjugacy of (2) on any interval, and by Theorem 4 a necessary condition for $[EWD]$. So tests for the validity of $[R]$ will indicate almost all there is to know about whether (2) is disconjugate on some interval.

$C(t)$ is assumed non-negative. The function $\psi(t)$ of (6) above can be defined with any initial value at any point, and $R[\psi(t)] \geq 0$. So a necessary and sufficient condition for $[R]$ on some interval is the existence of a differentiable function $Q(t)$ with $R[Q(t)] \leq 0$ on the interval, from Theorem 1 of Chapter 2.

Conversely, suppose on some open interval (a, b) there is a differentiable function $P(t)$ with $R[P(t)] \geq 0$, $P(t) \to +\infty$ as $t \to a$, and $P(t)$ is not bounded below in (a, b) . Then $[R]$ cannot hold in $[a, b]$ obviously, since at some point t near a , a solution $W(t)$ of (1) existing on $[a, b)$ has $W(t) < P(t)$, and less obviously, neither can $[R]$ hold in $(a, b]$ nor $[a, b)$. So $[WD]$ certainly does not hold in $[a, b]$ either, by Theorem 2. For applications a restriction to compact intervals is unimportant, since if a system is not disconjugate on an interval, then it is not disconjugate on some compact sub-interval.

The main test of the first kind, sufficient for disconjugacy, is the comparison test. If a solution $W_1(t)$ of

$$R_1\left[W_1\right] = W_1' + A_1(t) + B_1^*(t)W_1 + W_1B_1(t) + W_1C_1(t)W_1 \leq 0$$

exists on an interval J , and

$$R_2[W] = W' + A_2(t) + B_2^*(t)W + WB_2(t) + WC_2(t)W$$

and

$$\begin{pmatrix} A_1(t) & B_1^*(t) \\ B_1(t) & C_1(t) \end{pmatrix} \geq \begin{pmatrix} A_2(t) & B_2^*(t) \\ B_2(t) & C_2(t) \end{pmatrix}$$

on J , then

$$\begin{aligned} R_2[W_1(t)] &= W_1' + \begin{bmatrix} I & W_1 \end{bmatrix} \begin{bmatrix} A_2(t) & B_2^*(t) \\ B_2(t) & C_2(t) \end{bmatrix} \begin{bmatrix} I \\ W_1 \end{bmatrix} \\ &\leq W_1' + \begin{bmatrix} I & W_1 \end{bmatrix} \begin{bmatrix} A_1(t) & B_1^*(t) \\ B_1(t) & C_1(t) \end{bmatrix} \begin{bmatrix} I \\ W_1 \end{bmatrix} \\ &= R_1\left[W_1(t)\right] \leq 0 \end{aligned}$$

on J . So $[R]$ must hold on J .

A specially important case of an everywhere disconjugate system is one where $A(t) \leq 0$. Then $R[0] = A(t) \leq 0$, so $[R]$ holds everywhere.

A number of tests of the second kind are indicated at the end of the next chapter, which considers a series of solutions of Riccati inequalities. These solutions are bounds to solutions of the Riccati equation, and although successive bounds are more complicated, they are also better approximations. If the bounds converge to a solution, then the series of tests for non-existence of solutions of $R[W] = 0$ based on their behaviour is likely to be exhaustive.

Below, the most elementary of this series of tests is given. Despite its simplicity it seems to include as corollaries most explicit tests for oscillation so far produced.

THEOREM 5. *If a, b, c, d are four points for which $a \leq b \leq c \leq d$, $C(t) \geq 0$ on $[a, d]$, $A(t) \geq 0$ on $[a, b], [c, d]$, then if $R[W] = W' + A(t) + WC(t)W = 0$ has a solution on $[a, d]$ it is necessary that for some $e > 0$, and so for all e sufficiently large,*

$$\left\{e^{-1}I + \int_a^b C(u)du\right\}^{-1} - \int_b^c A(u)du + \left\{e^{-1}I + \int_c^d C(u)du\right\}^{-1} > 0 \, . \tag{29}$$

Proof. Suppose $W(t)$ is a solution of $R[W] = 0$ on $[a, d]$, and e is sufficiently large so that

$$eI > W(a) \, , \quad -eI < W(d) \, .$$

Let

$$U(t) = \left\{e^{-1}I + \int_a^t C(u)du\right\}^{-1}$$

on $[a, b]$. Then

$$\begin{aligned} R[U(t)] &= -U(t)C(t)U(t) + A(t) + U(t)C(t)U(t) \\ &= A(t) \geq 0 \end{aligned}$$

on $[a, b]$. $U(t)$ exists on $[a, b]$ and $U(a) = eI > W(a)$. So $U(t) > W(t)$ on $[a, b]$, by Theorem 1 of Chapter 2, and in particular, $U(b) > W(b)$.

Similarly

$$W(t) > \left\{-e^{-1}I - \int_t^d C(u)du\right\}^{-1}$$

on $[c, d]$. Let

$$U(t) = U(b) - \int_b^t A(u)du$$

on $[b, c]$. Then $R[U(t)] = U(t)C(t)U(t) \geq 0$ on $[b, c]$, $U(b) > W(b)$ so $U(t) > W(t)$ on $[b, c]$. Therefore,

$$\left\{e^{-1}I + \int_a^b C(u)du\right\}^{-1} - \int_b^c A(u)du = U(c) > W(c)$$

$$> \left\{-e^{-1}I - \int_c^d C(u)du\right\}^{-1} .$$

COROLLARY. *Suppose* $R[W] = 0$ *is defined and has a solution on* $[a, g)$, $d < g$, $A(t) \geq 0$ *in* $[c, g)$, *and let*

$$K(e, d) = \lim_{f \to g} \left\{eI + \int_d^f C(u)du\right\}^{-1} .$$

Let e' *be another positive number, and* $d' \in [c, g)$. *Then*

$$K(e, d) - K(e', d') = K(e, d)\left\{(e'-e)I + \int_d^{d'} C(u)du\right\}K(e', d') ,$$

and therefore both $K(e, d)$ *and* $K(e', d')$ *have the same nullspace* N . *If* $x \in N$, *and* $R[W] = 0$ *has a solution on* $[a, g)$, *then for some* $e > 0$,

$$x^*\left(eI + \int_a^b C(u)du\right)^{-1} x \geq x^* \int_b^c A(u)dux ,$$

since in the application of Theorem 5*, d can be taken arbitrarily close to g .*

In particular if N is the whole space, so $\int_d^f C(u)du \to \infty$ as $f \to g$, then $\int_b^c A(u)du \leq \left(eI + \int_a^b C(u)du\right)^{-1}$ for e sufficiently small and positive.

This is a result due to Ahlbrandt [*2*].

Alternatively suppose r eigenvalues of $\int_d^f C(u)du \to \infty$ as $f \to g$, so $\dim(N) = r$, and s eigenvalues of $\int_d^f A(u)du \to \infty$ as $f \to g$, and $r+s > n$. Then clearly $x^* \int_b^c A(u)du\, x$ cannot be bounded as c increases for all x in N , so $R[W] = 0$ cannot have a solution on $[a, g)$. This is a result due essentially to Tomastik [*1*].

Remark. If, in (29), $\int_a^b C(u)du$ and $\int_c^d C(u)du$ are invertible, then e can be assumed arbitrarily large, and so omitted. Also the requirement that $C(t) \geq 0$ on $[a, b]$ and $[c, d]$ can be considerably relaxed, since it is used here only to ensure the existence of $\left(e^{-1}I + \int_a^t C(u)du\right)^{-1}$ on $[a, b]$.

Example of the application of Theorem 5. Consider the Hamiltonian system

$$y' = z , \quad z' = -y , \tag{30}$$

and the corresponding Riccati equation

$$w' + 1 + w^2 = 0 , \tag{31}$$

for $t > 0$. Let $a = 0$, $b = 1$, $c = 3$. Then

$$\left(\int_0^1 C(u)du\right)^{-1} - \int_1^3 A(u)du = -1 \ .$$

So (31) has no solution on $[a, d]$ if $\left(\int_3^d C(u)du\right)^{-1} - 1 \leq 0$, that is, if $d \geq 4$.

So if $(\hat{y}, \hat{z})$ is the solution of (30) with $\hat{y}(0) = 0$, the next zero of $\hat{y}(t)$ occurs before $t = 4$.

Remark on applications of Theorem 5. To use the theorem to establish the absence of disconjugacy on some interval, or more usually, to put an upper bound to the distance of the next conjugate point, at least three points b, c, d have to be chosen successfully. The criterion for the choice is to ensure a rapid rate of descent of the upper bound function $U(t)$ for $t > a$. From an arbitrary point h , $U(t)$ can be defined in two ways (if both $A(t) \geq 0$ and $C(t) \geq 0$ for $t > h$) as follows:

$$U(t) = U(h) - \int_h^t A(u)du$$

or

$$U(t) = U(h)\left(I + \int_h^t C(u)duU(h)\right)^{-1} \ .$$

In the first approach $U'(h) = -A(h)$, and in the second

$$U'(h) = -U(h)C(h)U(h) \ .$$

So if $U(h)$ is close to zero, the first method of continuation will probably be more successful, and if $U(h)$ is bounded well away from zero, the second will be preferred.

In the above example the choice of continuation is optimal.

Summary

All the results obtained so far have been for intervals open at the left-hand end. Results for intervals open at the right-hand end can be obtained either by analogous proofs, or deduced by mapping t

onto $-t$ and solutions $W(t)$ of (1) onto $-W(-t)$. However, to clarify what is happening, and also to conveniently summarise the rather scattered results of this chapter, we re-express the principal results below for right-hand open intervals.

LEMMA 2. *If* $W_1(t)$ *is a solution of* $R[W] \le 0$ *existing in some interval* $[a, b)$ *, and* $W_2 \ge W_1(a)$ *, then the solution* $W_2(t)$ *of* (1) *with* $W_2(a) = W_2$ *exists on* $[a, b)$ *and* $W_2(t) \ge W_1(t)$ *on* $[a, b)$.

THEOREM 1. *Suppose* (1) *has a solution* $W_1(t)$ *existing on* $[a, b)$. *Then* (2) *has a principal solution* $\langle \hat{Y}(t), \hat{Z}(t) \rangle$ *in* $[a, b)$ *and for any solution* $W(t)$ *of* (1) *existing in* $[a, b)$ *,* $\hat{Y}^*(t)\hat{Z}(t) - \hat{Y}^*(t)W(t)\hat{Y}(t) \le 0$.

If $\langle Y(t), Z(t) \rangle$ *is a solution of* (2) *for which* $Y(t)$ *is invertible, and* $N = \hat{Y}^*(t)Z(t) - \hat{Z}^*(t)Y(t)$ *, then* $NY^{-1}(t)\hat{Y}(t) \to 0$ *as* $t \to b$.

COROLLARY 1. *Let* $[R]$ *hold on* $[a, b)$ *and* $\langle Y, Z \rangle$ *be a solution of* (2). *Then* $Y(t)$ *has a non-contracting nullspace iff, for all* λ *large enough,*

$$Y(a)\{Z(a)-\lambda Y(a)\}^{-1} \le \hat{Y}(a)\{\hat{Z}(a)-\lambda\hat{Y}(a)\}^{-1} \le 0 \ .$$

Then for $t \in [a, b)$ *,* $Y(t)\xi = 0$ *implies* $\{\hat{Z}^*(a)Y(a)-\hat{Y}^*(a)Z(a)\}\xi = 0$ *and* $\hat{Y}(t)\{\hat{Z}(a)-\lambda\hat{Y}(a)\}^{-1}\{Z(a)-\lambda Y(a)\}\xi = 0$.

So $Y(t)$ *is invertible in* $(a, b]$ *if*

$$Y(a)\{Z(a)-\lambda Y(a)\}^{-1} < \hat{Y}(a)\{\hat{Z}(a)-\lambda\hat{Y}(a)\}^{-1} \le 0$$

or if $Y^*(a)\hat{Z}(a) - Z^*(a)\hat{Y}(a)$ *is invertible and*

$$\hat{Y}^*(a)\hat{Z}(a) - \hat{Y}^*(a)Z(a)Y^{-1}(a)\hat{Y}(a) \le 0 \ .$$

COROLLARY 2. $[R]$ *on* $[a, b)$ *implies that the nullspace of* $\hat{Y}(t)$ *is a non-contracting set as* t *increases on* $[a, b)$.

COROLLARY 3. *If* $[a, b] \in I$ *, the principal solution* $\langle \hat{Y}, \hat{Z} \rangle$ *of* (2) *at* b *exists, and* $\hat{Y}(b) = 0$.

COROLLARY 4. $[R] \Rightarrow [WD]$ *on any interval.*

COROLLARY 5. *If* $[R]$ *holds on* $[a, b)$, $\hat{Y}(a)$ *is invertible iff* $P[a, b)$ *holds. Then a minimal solution* $\hat{W}(t)$ *exists iff* $[R]$ *and* $M_-(b)$ *hold.*

THEOREM 2. *On a compact interval* $[WD] \Rightarrow [R]$.

THEOREM 3. *If* $[WD]$ *holds on* $[a, b)$ *there is a sub-interval* $[c, b)$ *on which* $[R]$ *holds.*

LEMMA 4. *Let* $J = [a, c]$ *or* $[a, c)$. *If there is a non-degenerate solution* $\langle Y, Z \rangle$ *of* (2) *where* $Y(t)$ *has a non-contracting nullspace on* J , *then* $[R]$ *holds on* J .

THEOREM 4. *On any interval* $[R] \Longleftrightarrow [EWD]$.

THEOREM 5 is unchanged.

Appendix

In Chapter 1, it was shown that differential inequalities can be seen as general consequences of the properties of cones in vector spaces. In Chapter 2, we showed that, taking this geometric approach, the Riccati equation had a unique status with respect to symmetric matrix ordering. And in Chapter 3, we have used special properties of the Riccati equation to demonstrate the existence of maximal and minimal solutions on intervals.

From a geometrical point of view, it would be surprising if spaces of dimension $\frac{1}{2}n(n+1)$, that is, the spaces of symmetric matrices, had some special property or significance. The arguments in Chapter 3 ought to have analogues in spaces of any finite dimension.

The Lorentz ordering of Example 3, Chapter 1 was shown to imply that a function

$$f(x) = a\{x, x\} - 2x\{a, x\} + \alpha x + \{S, x\} + b \qquad \text{(A1)}$$

is of type K_+ and K_- , where a, b are n-vectors, α a scalar and S an $n \times n$ skew symmetric matrix, $\{\ \}$ the Lorentz product, and $\{S, x\}$ denotes the vector with components $\{s_i, x\}$, where s_i are the columns of S .

It is of interest that, just as the Riccati equation is unique among matrix equations in being of type K_+ and K_- , so (A1) is unique in being of type K_+ and K_- in Lorentz ordering, if $n \geq 3$. The proof uses essentially the same special case as does Theorem 2 of Chapter 2, namely the case when $n = 3$, and extends to higher dimensions in much the same way.

Theorem 1 of this chapter shows that given one solution of the Riccati equation, all solutions can be generated in terms of solutions of a linear equation involving the known solution, and furthermore when generated in this way, one solution, which can be explicitly nominated, is maximal. For equations of the type

$$\begin{aligned} x' &= f(t, x) \\ &= a(t)\{x, x\} - 2x\{a(t), x\} + \alpha(t)x + \{S(t), x\} + b(t) \qquad \text{(A2)} \end{aligned}$$

derived from (A1) with all coefficients continuous, a similar argument to that of Theorem 1 applies, as outlined below:

Let $y(t)$ be a solution of (A2) existing on an interval, and so transform (A2) by translation:

$$\begin{aligned} (x-y)' &= a\{(x-y), (x-y)\} - 2(x-y)\{a, x-y\} \\ &\quad + 2a\{x-y, y\} - 2(x-y)\{a, y\} - 2y\{a, x-y\} + \alpha(x-y) + \{S, x-y\} \qquad \text{(A3)} \\ &= f_y(x-y) \qquad \text{(A4)} \end{aligned}$$

where the argument t has been suppressed, and f_y has the same form as f, but with no constant term. Denoting the coefficients of f_y, a, S_1, α_1, and putting $u = x - y$, we have

$$u' = a\{u, u\} - 2u\{a, u\} + \{S_1, u\} + \alpha_1 u .$$

Now

$$\begin{aligned} \tfrac{1}{2}\{u, u\}' &= \{u, u'\} \\ &= -\{u, a\}\{u, u\} + \alpha_1\{u, u\} \quad \text{since} \quad \{u, \{S_1, u\}\} = 0 \qquad \text{(A5)} \end{aligned}$$

and $\left(\{u, u\}^{-1}\right)' = 2\left(\{u, a\}-\alpha_1\right)\{u, u\}^{-1}$. Therefore

$$\begin{aligned} \left(u\{u, u\}^{-1}\right)' &= \{u, u\}^{-1}\left(u'+2\left(\{u, a\}-\alpha_1\right)u\right) \\ &= a + \left\{S_1, u\{u, u\}^{-1}\right\} + \alpha_1 u\{u, u\}^{-1} . \qquad \text{(A6)} \end{aligned}$$

Let $v = u\{u, u\}^{-1}$, and $Y(t)$ be a fundamental matrix of the linear equation

$$z' = \{S_1(t),\ z\} + \alpha(t)z\ . \tag{A7}$$

Then

$$v(t) = v\left(t_0\right) + Y(t) \int_{t_0}^{t} Y^{-1}(u)a(u)du\ . \tag{A8}$$

And for each u, $w(t) = Y(t)Y^{-1}(u)a(u)$ is a solution of (A7) which is a special case of the type (A2) with $w(u) = a(u)$, and $z(t) \equiv 0$ is a solution of (A7). So if $a(u) \in C$, then $w(t) \in C$ by comparison with the solution $z(t) \equiv 0$, and if $a(u) \in C^{i}$, $w(t) \in C^{i}$, where C^{i} is the interior of C. Then $v(t)$ is increasing if $a(u) \in C^{i}$ for all u, and the solution we would like to fulfil the role of principal solution on $(c,\ b]$ say is $\hat{v}(t) = Y(t) \int_{c}^{t} Y^{-1}(u)a(u)du$. The proof that the indefinite integral exists is like that of Theorem 1; that is, there are solutions $v(t)$ of (A6) with $v(t) \geq 0$ on $(c,\ b]$, and $v(t)$ is non-increasing, (if $a(t) \in C$) so $Y(b) \int_{t}^{b} Y^{-1}(u)a(u)du \leq v(b)$ as $t \to c$, and so $Y(b) \int_{c}^{b} Y^{-1}(u)a(u)du$ exists.

In this development, the vector $s = u\{u,\ u\}^{-1}$ plays the role of an inverse, and then $s\{s,\ s\}^{-1} = u$. It can be verified that this "inversion" inverts ordering in C. So solutions of the first order equation (A6) correspond by simple transformations to solutions of (A2); in particular $\hat{v}(t)$ corresponds to a maximal solution on $(a,\ b]$: $\hat{x}(t) = y(t) + \hat{V}(t)\{\hat{V}(t),\ \hat{V}(t)\}^{-1}$. Verification that the solution is indeed maximal on $(a,\ b]$ can be achieved by essentially the same method as in Theorem 1.

Notes

The concept of conjugate points for systems originated in the Jacobi necessary condition in the calculus of variations (Morse [2], Bliss [1], Radon [1]). The controllability $[C]$ condition comes from the normality condition of the calculus of variations.

In the scalar case, the significance of Riccati equations has long been recognised (Bôcher [1], [2]). Principal solutions for systems are dealt with in Reid [4], [9], [14], Hartman [1], [2], and Coppel [3]. Reid and Coppel make extensive use of Riccati equations; Hartman used the reduced form that appears in our Theorem 1. All these writers assume a condition equivalent to our $[C]$ when defining principal solutions although Reid [15] defines, and proves the existence of, a distinguished solution (that is, maximal or minimal) of (1) under conditions like our M_+ and M_-, and under another condition of intermediate strength, which still gives an invertible principal solution.

The essential difference between our approach and previous approaches is that we concentrate first on the Riccati equation, for which the existence of a maximal or minimal solution, at least in some transformed domain, arises naturally without the need for any controllability or normality conditions, provided that there is at least one solution on the interval. Having regarded $[R]$ as the primary condition, we then relate $[R]$ to disconjugacy.

For the transformation to unitary matrices or method of polar co-ordinates, which completely avoids problems of existence of solutions, see Atkinson [1], or Lidskii [1], or Coppel [1]. In our context, if $\langle Y, Z\rangle$ is a solution of (2), and

$$L(t) = \big(Z(t)+iY(t)\big)\big(Z(t)-iY(t)\big)^{-1}$$

then $L(t)$ is a unitary matrix, existing everywhere, and

$$2L' + -i(L-I)A(t)(L-I) + (L+I)B(t)(L-I) + (L-I)B^*(t)(L+I) + i(L+I)C(t)(L+I) = 0 .$$

As mentioned earlier, in this approach, order relations between solutions are difficult to express; in fact, the simplest way is to transform back into a symmetric matrix domain, the nett result being equivalent to our translations and inversions.

A clear exposition of the properties of the matrix Riccati equation, including the unsymmetric equation, is given in Barnett's recent book [1, Chapter 5], which also contains a useful chapter on generalized inverses. The paper of Levin [1] is also useful.

CHAPTER 4

CONTINUED FRACTION EXPANSIONS OF SOLUTIONS OF THE RICCATI EQUATION

Introduction and summary

In this chapter we consider a solution of the Riccati equation

$$R[W] = W' - A(t) + WC(t)W = 0 , \qquad (1)$$

with $W(t_0) = S_0$.

$A(t)$ and $C(t)$ are symmetric continuous $n \times n$ matrix functions, and are usually each considered to be either positive or negative definite, and defined for all t .

If a sequence $Z_n(t)$ is defined recursively by

$$Z_1(t) = S_0 + \int_{t_0}^{t} A(u)du ,$$

$$Z_n(t) = Z_{n-1}^{-1}(t) \ldots Z_1^{-1}(t) \int_{t_0}^{t} Z_1(u) \ldots Z_{n-1}(u)B_n(u)Z_{n-1}(u) \ldots$$
$$\ldots Z_1(u)duZ_1^{-1}(t) \ldots Z_{n-1}^{-1}(t)$$

where

$$B_n(t) = \begin{cases} A(t) & \text{if } n \text{ is odd,} \\ C(t) & \text{if } n \text{ is even,} \end{cases}$$

then we define a continued fraction associated with $W(t)$ by

$$\left(Z_1^{-1}(t)+\left(Z_2^{-1}(t)+\left(Z_3^{-1}(t) + \ldots\right)^{-1}\right)^{-1}\right)^{-1} .$$

By this is meant, in effect, a sequence of convergents $T_1(t) = Z_1(t)$, $T_2(t) = \left(Z_1^{-1}(t)+Z_2(t)\right)^{-1}$, $T_3(t) = \left(Z_1^{-1}(t)+\left(Z_2^{-1}(t)+Z_3(t)\right)^{-1}\right)^{-1}$ etc. Each convergent $T_n(t)$ satisfies a Riccati equation $R\{T_n(t)\} = (-1)^n K_n(t)$, where $K_n(t)$, whose nature and existence is established in Theorem 1, has a sign depending on $B_{n-1}(t)$, that is alternatively on the sign of $A(t)$ and $C(t)$, if these are of fixed sign on $[t_0, t]$.

There is a digression in Lemma 4 to show that matrix continued fractions share with their scalar counterparts the property that their convergents can be expressed as the ratio of two linearly independent solutions of a second order linear recurrence relations. Otherwise Lemmas 2-6 are concerned with determining local behaviour near t_0 of the convergents $T_n(t)$ in relation to each other and to the solution $W(t)$ of (1). This is done firstly so the Riccati inequalities of Theorem 1 can be applied to show that the sequences $\{T_{2n}(t)\}$ and $\{T_{2n+1}(t)\}$ form a sequence of bounds to $W(t)$, (Theorem 2); whether the bounds are upper or lower depends on the sign of $A(t)$ and $C(t)$ respectively, but in any case they improve as n increases (Theorem 4). The second reason for looking at local behaviour is that it shows that $\{T_n(t)\}$ is a sequence of rapidly improving approximants to $W(t)$; in fact

$$W(t) - T_n(t) = O\left(|t-t_0|^{2n-1}\right), \quad \text{(Theorem 3).}$$

From the inverse equation $V' = C(t) - VA(t)V$ another continued fraction can be derived. This is important because in this way bounds to the maximal (or minimal) solution of the previous chapter can be obtained. If those bounds fail to exist on an interval, then so does the principal* solution they approximate (Theorem 5). Therefore a pair of conjugate points exists on the interval, and the associated Hamiltonian system is not disconjugate. Many existing oscillation criteria are derivable as applications of this principle.

Unfortunately any proof that the continued fractions we derive in fact converge to a solution of (1) presents formidable difficulties, although it is easy enough in many applications. Often the continued fraction converges almost everywhere; for example

* In this and the following chapter, the maximal and minimal solutions of (1) on an interval are often referred to as principal solutions of (1) at the appropriate endpoint.

$$\tan t = \cfrac{t}{1 - \cfrac{t^2}{3 - \cfrac{t^2}{5} \dots}} ,$$

converges for all t (real or complex) except $t = \frac{\pi}{2} \pm n\pi$.

The continued fraction expansion is important for disconjugacy theory because it offers a series of necessary conditions for disconjugacy expressed in terms of the coefficient functions. If the continued fraction does converge to the desired solution, then satisfaction of the series of necessary conditions is a sufficient condition.

Because so many terms and symbols are introduced, an index of definitions is appended to this chapter.

The basic equation satisfied by the convergents

DEFINITIONS. *Let* $\{V_n(t)\}$, $\{C_n(t)\}$ *and* $\{A_n(t)\}$ *be sequences of symmetric continuous matrix functions defined, for* $t \neq t_0$ *, by the relations*

$$C_0(t) = C(t) \ , \quad A_0(t) = A(t) \ , \quad V_1(t) = S_0 + \int_{t_0}^{t} A(u)du \ ,$$

where S_0 *is a symmetric matrix, and*

$$C_n(t) = V_n^{-1}(t)A_{n-1}(t)V_n^{-1}(t) \ , \quad n = 1, 2, \dots \tag{2}$$

$$A_n(t) = V_n(t)C_{n-1}(t)V_n(t) \ , \quad n = 1, 2, \dots \tag{3}$$

$$V_{n+1}(t) = \int_{t_0}^{t} A_n(u)du \ , \quad n = 1, 2, \dots \tag{4}$$

The sequences can be assumed to terminate where any definition fails due to $V_n(t)$ failing to be invertible. However we shall usually assume $A(t)$ and $C(t)$ to each be of definite sign everywhere, and so, provided $V_1(t)$ is invertible on some interval including t_0 then all the other terms of all the sequences will exist and be either positive definite or negative definite on the interval.

Whenever in what follows we assert that a matrix function, whose definition involves $V_n(t)$ for some n, exists, it will be understood that for $i = 1 \dots n-1$, $V_i(t)$ exists and is invertible.

The simplest way to see why there should be a continued fraction solution of (1) is as follows. Let $\{S_n(t)\}$ be another sequence of differentiable symmetric matrix functions, with $S_0(t)$ the solution $S(t)$ of (1) with $S(t_0) = S_0$, and

$$S_n(t) = V_n(t)S_{n-1}^{-1}(t)V_n(t) - V_n(t) .$$

Then $S_n(t)$ satisfies $S_n'(t) = A_n(t) - S_n(t)C_n(t)S_n(t)$. And near t_0, $V_1(t) \to S_0$, $V_2(t) = O(|t-t_0|)$, and if S_0 is invertible, $A_2(t) = O\left(|t-t_0|^2\right)$, etc. $A_n(t)$ becomes small as n increases, and so does $S_n(t)$, at least near t_0.

By putting $S_n(t) \equiv 0$ for some n, and solving the n equations defining $S_n(t)$ to recover the approximated value of $S_0(t)$, we generate a sequence $\{T_n(t)\}$:

$$T_1(t) = V_1(t) ,$$

$$T_2(t) = V_1(t)\left(V_1(t)+V_2(t)\right)^{-1}V_1(t) ,$$

$$T_3(t) = V_1(t)\left(V_1(t)+V_2(t)\left(V_2(t)+V_3(t)\right)^{-1}V_2(t)\right)^{-1}V_1(t) ,$$

etc. The somewhat neater formulation of the $\{T_n(t)\}$ given in the introduction will be arrived at later.

The next set of sequences are introduced for the rather extensive manipulations required to show that the $\{T_n(t)\}$ are solutions of a Riccati inequality (and equation) as in (9) below. Later (Lemma 7) we will be able to produce a neater and more natural version of (9).

DEFINITIONS. *Let* $\{X_{n,r}(t)\}$ *be a double sequence of symmetric matrices defined for* $r \le n$, *by*

$$X_{n,0}(t) = 0 \text{ ,}$$

$$X_{n,r+1}(t) = V_{n-r}(t)\left(X_{n,r}(t)+V_{n-r}(t)\right)^{-1}V_{n-r}(t) \text{ ,} \quad 0 \le r \le n \text{ .} \tag{5}$$

Then $X_{n,n}(t) = T_n(t)$.

Let

$$H_{n,r}(t) = V_r(t)\left(X_{n,n-r}(t)+V_r(t)\right)^{-1} = X_{n,n-r+1}(t)V_r^{-1}(t) \tag{6}$$

and let $M_{n,r}(t) = M_{n,r-1}(t)H_{n,r}(t)$, $M_{n,0}(t) = I$, *so*

$$M_{n,r}(t) = H_{n,1}(t) \ldots H_{n,r}(t) \text{ .} \tag{7}$$

Let

$$K_{n,r}(t) = (-1)^r M_{n,r}(t)A_r(t)M^*_{n,r}(t) \text{ ,}$$

$$r = 0, \ldots, n \text{ ,} \quad K_{n,n+1}(t) = 0 \text{ .} \tag{8}$$

Again we assume that $\{X_{n,r}(u)\}$, $\{H_{n,r}(u)\}$, $\{M_{n,r}(u)\}$ and $\{K_{n,r}(u)\}$ cease to be defined beyond any n where their definition involves the inverse of a singular matrix.

All sequences are certainly well-defined for all n if $A(t)$ and $C(t)$ are both positive definite (or both negative definite) on $[t_0, t]$, and S_0 non-negative (or, respectively, non-positive), since then $V_r(t) > 0$ if $t > t_0$, and $X_{n,r}(t) > 0$ for $r = 1, \ldots, n$, $t > t_0$, and $X_{n,r}(t) + V_{n-r}(t)$ is always invertible.

THEOREM 1. *If the sequences* $\{X_{n,r}(t)\}$, $\{H_{n,r}(t)\}$, $\{M_{n,r}(t)\}$ *and* $\{K_{n,r}(t)\}$ *are well-defined, then*

$$R\{T_n\} = -K_{n,n}(t) \text{ ,} \quad t \ne t_0 \text{ .} \tag{9}$$

Proof. We omit the suffix n , and the argument t , for the time being. Now

$$H_n = V_n\left(X_{n-n}+V_n\right)^{-1} = I$$

and

$$H_r = V_r\left(X_{n+1-(r+1)}+V_r\right)^{-1} = V_r\left(H_{r+1}V_{r+1}+V_r\right)^{-1} .$$

So

$$H_rH_{r+1}V_{r+1} = \left(I-H_r\right)V_r . \tag{10}$$

And

$$\begin{aligned}\left(H_r-I\right)A_{r-1}\left(H_r^*-I\right) &= H_rH_{r+1}V_{r+1}V_r^{-1}A_{r-1}V_r^{-1}V_{r+1}H_{r+1}^*H_r^* \\ &= H_rH_{r+1}V_{r+1}A_rV_{r+1}H_{r+1}^*H_r^* \\ &= H_rH_{r+1}A_{r+1}H_{r+1}^*H_r^* .\end{aligned}$$

Then

$$\begin{aligned}M_{r-1}\left(I-H_r\right)A_{r-1}\left(I-H_r^*\right)M_{r-1}^* &= M_{r-1}H_rH_{r+1}A_{r+1}H_{r+1}^*H_r^*M_{r-1}^* \\ &= M_{r+1}A_{r+1}M_{r+1}^* = (-1)^{r+1}K_{r+1} \quad \text{if} \quad 0 \le r < n .\end{aligned}$$

And $M_{n-1}\left(I-H_n\right)A_{n-1}\left(I-H_n^*\right)M_{n-1}^* = 0 = (-1)^{n+1}K_{n+1}$ since $H_n = I$.

Let $L_r = (-1)^rM_rX'_{n-r}M_r^*$, $L_n = 0$. Then for $1 \le r \le n$,

$$\begin{aligned}X'_{n-r+1} &= V'_r\left(X_{n-r}+V_r\right)^{-1}V_r + V_r\left(X_{n-r}+V_r\right)^{-1}V'_r \\ &\qquad - V_r\left(X_{n-r}+V_r\right)^{-1}\left(X'_{n-r}+V'_r\right)\left(X_{n-r}+V_r\right)^{-1}V_r \\ &= H_rV'_r + V'_rH_r^* - H_rV'_rH_r^* - H_rX'_{n-r}H_r^* \\ &= -\left(H_r-I\right)V'_r\left(H_r^*-I\right) + V'_r - H_rX'_{n-r}H_r^* .\end{aligned}$$

Therefore,

$$\begin{aligned}(-1)^{r-1}M_{r-1}X'_{n-r+1}M_{r-1}^* &= (-1)^rM_{r-1}\left(H_r-I\right)A_{r-1}\left(H_r^*-I\right)M_{r-1}^* \\ &\qquad + (-1)^{r-1}M_{r-1}A_{r-1}M_{r-1}^* + (-1)^rM_rX'_{n-r}M_r^*\end{aligned}$$

since $V'_r = A_{r-1}$. Therefore, $L_{r-1} = K_{r-1} + (-1)^r(-1)^{r+1}K_{r+1} + L_r$.

Therefore,

$$L_0 = \sum_{r=1}^{n} \left(L_{r-1}-L_r\right) \quad \text{since} \quad L_n = 0$$

$$= \sum_{r=1}^{n} \left(K_{r-1}-K_{r+1}\right) = K_0 + K_1 - K_n \quad \text{since} \quad K_{n+1} = 0 \; .$$

But $L_0 = X'_n$ since $M_0 = I$, and $K_0 = A$,

$$K_1 = -H_1A_1H_1^* = -X_nV_1^{-1}V_1CV_1V_1^{-1}X_n$$
$$= -X_nCX_n \; .$$

So $R\left[X_n\right] = X'_n - A + X_nCX_n = -K_n$.

On replacing the omitted suffices, and with $T_n(t) = X_{n,n}(t)$,

$$R\left[T_n\right] = -K_{n,n}(t) \; . \tag{9}$$

Note that $K_{n,n}(t)$ has the same sign as $A(t)$ if n is even, or $-C(t)$ if n is odd, assuming $A(t)$ or $C(t)$ respectively are sign definite.

Behaviour of convergents and associated functions near t_0

DEFINITIONS. *Let*

$$Z_n(t) = \begin{cases} V_1(t)V_2^{-1}(t) \;\ldots\; V_{n-1}^{-1}(t)V_n(t)V_{n-1}^{-1}(t) \;\ldots\; V_1(t) & \textit{if } n \textit{ is odd,} \\ \\ V_1^{-1}(t)V_2\;(t) \;\ldots\; V_{n-1}^{-1}(t)V_n(t) \;\ldots\; V_1^{-1}(t) & \textit{if } n \textit{ is even.} \end{cases} \tag{11}$$

Then it is easily verified that

$$V_n(t) = Z_1(t) \;\ldots\; Z_{n-1}(t)Z_n(t)Z_{n-1}(t) \;\ldots\; Z_1(t) \; . \tag{12}$$

Let

$$B_n(t) = \begin{cases} A(t) & \textit{if } n \textit{ is odd,} \\ \\ C(t) & \textit{if } n \textit{ is even.} \end{cases} \tag{13}$$

Until further notice $B_n(t)$ is assumed to be either positive or negative definite for all t and each n .

Let

$$\nu_n = \begin{cases} +1 & \text{if } B_n(t) > 0 , \\ -1 & \text{if } B_n(t) < 0 , \end{cases} \tag{14}$$

Let $\sigma(t) = \operatorname{sgn}(t-t_0)$.

Now

$$A_n(t) = V_n(t)V_{n-1}^{-1}(t)A_{n-2}(t)V_{n-1}^{-1}(t)V_n(t)$$

and $A_0(t) = A(t)$, $A_1(t) = V_1(t)C(t)V_1(t)$. So $\nu_n A_n(t) > 0$ and

$$\sigma(t)\nu_n V_n(t) = \int_{t_0}^{t} A_n(u)du > 0 \ , \quad n = 2, 3, \ldots \quad \text{if} \quad t \neq t_0 \ .$$

Therefore, $\sigma(t)\nu_n Z_n(t) > 0$ also, if $t \neq t_0$.

The sequence $\{Z_n(t)\}$ is introduced partly to give a simpler expression for the continued fraction associated with a solution of (1), and partly because its behaviour near t_0 is more amenable to investigation.

LEMMA 1. *If* A, B *are invertible, and* $|A^{-1}||B| < 1$, *then*

$$|(A+B)^{-1}| \leq \frac{|A^{-1}|}{1-|A^{-1}||B|} \ .$$

Proof.

$$\begin{aligned}(A+B)^{-1} &= A^{-1}(A+B-B)(A+B)^{-1} \\ &= A^{-1} - A^{-1}B(A+B)^{-1} \ .\end{aligned}$$

Therefore,

$$|(A+B)^{-1}| \leq |A^{-1}| + |A^{-1}||B||(A+B)^{-1}| \ .$$

Therefore,

$$|(A+B)^{-1}| \leq \frac{|A^{-1}|}{1-|A^{-1}||B|} \ .$$

The next lemma is needed to deal with the difficult case arising when S_0 is singular, but not zero.

LEMMA 2. *If* $|t-t_0|$ *is sufficiently small, and if on a range* J *including* t_0 *in its interior,* $0 < \alpha I \leq A(t)$, $C(t) \leq \beta I$, *then*

$$-\frac{8\beta I}{\alpha^2} \leq (t-t_0)\left(S_0 + \int_{t_0}^{t} A(u)du\right)^{-1} \leq \frac{8\beta I}{\alpha^2} , \ t \neq t_0 .$$

Proof. If $S_0 \geq 0$, $t > t_0$, then

$$\begin{aligned}\left(S_0 + \int_{t_0}^{t} A(u)du\right)^{-1} &\leq \left(\int_{t_0}^{t} A(u)du\right)^{-1} \\ &\leq \frac{I}{\alpha(t-t_0)} \leq \frac{8\beta}{\alpha}\,\frac{I}{\alpha(t-t_0)} .\end{aligned}$$

If $S_0 \ngeq 0$, let λ be its maximum negative eigenvalue. Suppose

$$(t-t_0) \leq \min\left\{\frac{-\lambda}{2\alpha}, \frac{-3\lambda}{8(\beta-\frac{1}{2}\alpha)}\right\} .$$

Then $S_0 + \frac{1}{2}\alpha(t-t_0)I$ has no eigenvalues between $\frac{1}{2}\alpha(t-t_0)$ and $\frac{1}{2}\alpha(t-t_0) + \lambda$, and the latter value $\leq \lambda - \frac{\lambda}{4} = \frac{3}{4}\lambda < 0$. Let $\varepsilon = \frac{1}{2}\alpha(t-t_0)$. Then $S_0 + \varepsilon I$ is invertible, and

$$(S_0+\varepsilon I)^{-1} \leq I\varepsilon^{-1} , \quad (S+I\varepsilon)^{-1} \geq I(\lambda+\varepsilon)^{-1} \geq \frac{4\lambda^{-1}}{3} I .$$

Further,

$$\int_{t_0}^{t} A(u)du - \varepsilon I \geq \tfrac{1}{2}\alpha(t-t_0)I = \varepsilon I > 0 .$$

So

$$\begin{aligned}\left(S_0 + \int_{t_0}^{t} A(u)du\right)^{-1} = (S_0+\varepsilon I)^{-1}\left((S_0+\varepsilon I)^{-1}+\left(\int_{t_0}^{t} A(u)du-\varepsilon I\right)^{-1}\right)^{-1} \\ \times \left(\int_{t_0}^{t} A(u)du-\varepsilon I\right)^{-1} .\end{aligned}$$

And

$$\left(S_0+\varepsilon I\right)^{-1} + \left(\int_{t_0}^{t} A(u)du-\varepsilon I\right)^{-1} \ge \frac{4\lambda^{-1}}{3} I + (\beta-\tfrac{1}{2}\alpha)^{-1}\left(t-t_0\right)^{-1}I$$

$$= (\beta-\tfrac{1}{2}\alpha)^{-1}\left(t-t_0\right)^{-1}I\left[1 + (\beta-\tfrac{1}{2}\alpha)\ \frac{4}{3\lambda}\ \left(t-t_0\right)\right]$$

$$\ge \tfrac{1}{2}(\beta-\tfrac{1}{2}\alpha)^{-1}\left(t-t_0\right)^{-1}I .$$

Therefore,

$$\left|\left(S_0 + \int_{t_0}^{t} A(u)du\right)^{-1}\right| \le \left|\left(S_0+\varepsilon I\right)^{-1}\right|\left|\left(S_0+\varepsilon I\right)^{-1} + \left(\int_{t_0}^{t} A(u)du-\varepsilon I\right)^{-1}\right|$$

$$\times \left|\left(\int_{t_0}^{t} A(u)du-\varepsilon I\right)^{-1}\right|$$

$$\le \left|\tfrac{1}{2}\alpha\left(t-t_0\right)\right|^{-1}2(\beta-\tfrac{1}{2}\alpha)\left|t-t_0\right|\left|\tfrac{1}{2}\alpha\left(t-t_0\right)\right|^{-1}$$

$$= \frac{8(\beta-\tfrac{1}{2}\alpha)}{\alpha^2}\ \left|t-t_0\right|^{-1} .$$

Therefore,

$$-\left|t-t_0\right|^{-1}\ \frac{8\beta}{\alpha^2}\ I \le \left(S_0 + \int_{t_0}^{t} A(u)du\right)^{-1} \le \left|t-t_0\right|^{-1}\ \frac{8\beta}{2}\ I$$

LEMMA 3. *There exist sequences* $\{\alpha_n\}$, $\{\beta_n\}$ *of bounds so that on some compact interval* J *including* t_0 ,

$$|Z_n(t)| \le \beta_n|t-t_0| , \qquad n = 2 \dots ,$$

$$\left|\left(Z_n(t)\right)^{-1}\right| \le \alpha_n|t-t_0|^{-1} , \quad n = 2 \dots , \tag{15}$$

and $|Z_1(t)| \le \beta_1$, $\left|\left(Z_1(t)\right)^{-1}\right| \le \alpha_1|t-t_0|^{-1}$.

Proof. Let α, β be bounds on $A(t)$, $C(t)$: $0 < \alpha I \le \nu_n B_n(t) \le \beta I$. Then

$$Z_1(t) = S_0 + \int_{t_0}^{t} A(u)du$$

is bounded on J . And $\left|Z_1^{-1}(t)\right| \le \alpha_1|t-t_0|^{-1}$ by Lemma 2, and

$$Z_1^{-1}(u)Z_1(t) = \left(S_0 + \int_{t_0}^{u} A(w)dw\right)^{-1}\left(S_0 + \int_{t_0}^{t} A(w)dw\right)$$

$$= I + Z_1^{-1}(u)\int_{u}^{t} A(w)dw \ .$$

Therefore,

$$\begin{aligned}\left|Z_1^{-1}(u)Z_1(t)\right| &\le 1 + \alpha_1|u-t_0|^{-1}\alpha|t-u| \\ &\le 1 + \alpha_1\alpha|u-t_0|^{-1}\left(|t-t_0|+|u-t_0|\right) \\ &= \gamma_1 + \gamma_2\frac{|t-t_0|}{|u-t_0|} \quad \text{for constants } \gamma_1, \gamma_2 \ . \qquad (16)\end{aligned}$$

Suppose it is true that for $m = 2 \ldots n-1$, the inequalities (15) hold, and $n > 1$. Let

$$\xi_{n-1}(u) = Z_{n-1}(u) \ldots Z_1(u)Z_1^{-1}(t) \ldots Z_{n-1}^{-1}(t)\xi$$

for an arbitrary vector ξ .

Then

$$\xi^*Z_n(t)\xi = \xi^*Z_{n-1}^{-1}(t) \ldots Z_1^{-1}(t)V_n(t)Z_1^{-1}(t) \ldots Z_{n-1}^{-1}(t)\xi \ .$$

And

$$V_n(t) = \int_{t_0}^{t} V_{n-1}(u)V_{n-2}^{-1}(u) \ldots V_1^{(-1)^n}(u)B_n(u) \ldots V_{n-1}(u)du$$

$$= \int_{t_0}^{t} Z_1(u) \ldots Z_{n-1}(u)B_n(u)Z_{n-1}(u) \ldots Z_1(u)du \ .$$

Therefore,

$$\xi^*Z(t)\xi = \int_{t_0}^{t} \xi_{n-1}^*(u)B_n(u)\xi_{n-1}(u)du \ ;$$

and

$$|\xi_{n-1}(u)| \le |Z_{n-1}(u)| \dots \left|Z_1(u)Z_1^{-1}(t)\right| \dots \left|Z_{n-1}^{-1}(t)\right| |\xi|$$

$$\le \beta_{n-1}\alpha_{n-1} \dots \beta_2\alpha_2 \left|\frac{u-t_0}{t-t_0}\right|^{n-1} \left(\gamma_1 + \gamma_2 \frac{|u-t_0|}{|t-t_0|}\right) |\xi| \ .$$

So,

$$|\xi^* Z(t)\xi| \le \left|\beta \int_{t_0}^{t} |\xi_{n-1}(u)|^2 du\right|$$

$$\le |\beta_{n-1}\alpha_{n-1} \dots \beta_2\alpha_2|^2 \beta |\xi|^2 \left|\int_{t_0}^{t} \left(\gamma_1 \left|\frac{u-t_0}{t-t_0}\right|^{n-1} + \gamma_2 \left|\frac{u-t_0}{t-t_0}\right|^{n}\right)^2 du\right|$$

$$\le |\beta_{n-1} \dots \alpha_2|^2 \beta |\xi|^2 |t-t_0| \left|\frac{\gamma_1^2}{2n-1} + \frac{2\gamma_2\gamma_1}{2n} + \frac{\gamma_2^2}{2n+1}\right|$$

$$= \beta_n |\xi|^2 |t-t_0| \ .$$

And

$$\xi = Z_{n-1}(t) \dots Z_1(t)Z_1^{-1}(u) \dots Z_{n-1}^{-1}(u)\xi_{n-1}(u) \ ,$$

so

$$|\xi| \le |Z_{n-1}(t)| \dots \left|Z_1(t)Z_1^{-1}(u)\right| \dots \left|Z_{n-1}^{-1}(u)\right| |\xi_{n-1}(u)|$$

$$= (\alpha_{n-1}\beta_{n-1} \dots \alpha_2\beta_2)\left(\gamma_1 + \gamma_2 \frac{|t-t_0|}{|u-t_0|}\right) \left|\frac{t-t_0}{u-t_0}\right|^{n-1} |\xi_{n-1}(u)| \ ,$$

and

$$\gamma_1 + \gamma_2 \frac{|t-t_0|}{|u-t_0|} \ge \gamma_3 \left|\frac{t-t_0}{u-t_0}\right|$$

if $|u-t_0| \le |t-t_0|$, for some $\gamma_3 > 0$. Therefore,

$$|\xi_{n-1}(u)| \ge (\alpha_{n-1} \dots \beta_2)^{-1}\gamma_3^{-1} \left|\frac{u-t_0}{t-t_0}\right| \ |\xi| \ .$$

Therefore,

$$\nu_n\sigma(t)\xi^*Z(t)\xi \geq \alpha\sigma(t)\int_{t_0}^{t} |\xi_{n-1}(u)|^2 du$$
$$\geq \sigma(t)|\xi|^2\alpha_n^{-1}|t-t_0|$$

where

$$\alpha_n^{-1} = \left(\alpha_{n-1}\beta_{n-1} \cdots \alpha_2\beta_2\right)^{-2} \frac{\alpha}{2n+1}\gamma_3^{-2} .$$

But ξ is arbitrary, so $|Z(t)| \leq \beta_n|t-t_0|$ and

$$\nu_n\sigma(t)Z(t) \geq \alpha_n^{-1}|t-t_0|I > 0 .$$

Therefore,

$$0 < \nu_n\sigma(t)Z^{-1}(t) \leq \alpha_n|t-t_0|^{-1}I$$

or $|Z^{-1}(t)| \leq \alpha_n|t-t_0|^{-1}$.

Just as the convergents of ordinary continued fractions can be expressed as the ratio of two solutions of a second order linear recurrence relation, so can the convergents of a matrix continued fraction. Furthermore, the solutions of the recurrence relations have a direct role in the approximating Riccati equations (for example (9)) of which the convergents are solutions. The results in the following lemma are general facts true for symmetric matrix continued fractions.

LEMMA 4.

$$T_n(t) = F_n(t)G_n^{-1}(t) \quad \textit{for} \quad n \geq 1 , \tag{17}$$

where $F_{-1}(t) = I$ *,* $F_0(t) = 0$ *,* $G_{-1}(t) = 0$ *,* $G_0(t) = I$ *and*

$$F_n(t) = F_{n-1}(t)Z_n^{-1}(t) + F_{n-2}(t) ,$$

$$G_n(t) = G_{n-1}(t)Z_n^{-1}(t) + G_{n-2}(t) ; \tag{18}$$

and

$$T_n(t) - T_{n-1}(t) = (-1)^{n-1}\left(G_n(t)G_{n-1}^*(t)\right)^{-1} . \tag{19}$$

Proof.

$$T_n(t) = \left(Z_1^{-1}(t)+\left(Z_2^{-1}(t) + \ldots + \left(Z_{n-1}^{-1}(t)+Z_n(t)\right)^{-1}\right)^{-1} \ldots\right)^{-1} .$$

Suppose $U_n(t) = T_n^{-1}(t) - Z_1^{-1}(t)$, so $T_n(t) = \left(Z_1^{-1}(t)+U_n(t)\right)^{-1}$.

Then $U_n(t)$ is a continued fraction partial approximant also.

Suppose that all approximants of order less than n can be formed by a second order recurrence relation, as in the hypothesis. Then $U_n(t) = P_{n-1}(t)Q_{n-1}^{-1}(t)$ where

$$P_m(t) = P_{m-1}(t)Z_{m+1}^{-1}(t) + P_{m-2}(t) ,$$

$$Q_m(t) = Q_{m-1}(t)Z_{m+1}^{-1}(t) + Q_{m-2}(t) \quad \text{for } m = 1 \ldots n-1 ,$$

$$P_{-1}(t) = Q_0(t) = I , \quad P_0(t) = Q_{-1}(t) = 0 .$$

And

$$\begin{aligned} P_{n-1}(t)Q_{n-1}^{-1}(t) = U_n(t) &= G_n(t)F_n^{-1}(t) - Z_1^{-1}(t) \\ &= \left(G_n(t)-Z_1^{-1}(t)F_n(t)\right)F_n^{-1}(t) \end{aligned}$$

so we can take

$$P_{n-1}(t) = G_n(t) - Z_1^{-1}(t)F_n(t) ,$$

$$Q_{n-1}(t) = F_n(t) .$$

Then

$$\begin{aligned} F_n(t) = Q_{n-1}(t) &= Q_{n-2}(t)Z_n^{-1}(t) + Q_{n-3}(t) \\ &= F_{n-1}(t)Z_n^{-1}(t) + F_{n-2}(t) . \end{aligned}$$

And

$$\begin{aligned} G_n(t) &= P_{n-1}(t) + Z_1^{-1}(t)Q_{n-1}(t) \\ &= P_{n-2}(t)Z_n^{-1}(t) + P_{n-3}(t) + Z_1^{-1}(t)\left[Q_{n-2}(t)Z_n^{-1}(t)+Q_{n-3}(t)\right] \\ &= \left[P_{n-2}(t)+Z_1^{-1}(t)Q_{n-2}(t)\right]Z_n^{-1}(t) + P_{n-3}(t) + Z_1^{-1}(t)Q_{n-3}(t) \\ &= G_{n-1}(t)Z_n^{-1}(t) + G_{n-2}(t) \ . \end{aligned}$$

But the formula is easily verified when $n = 1$, so by induction it is true for all n .

Then

$$\begin{aligned} F_n(t)G^*_{n-1}(t) &= F_{n-1}(t)Z_n^{-1}(t)G^*_{n-1}(t) + F_{n-2}(t)G^*_{n-1}(t) \\ &= F_{n-1}(t)\left(G^*_n(t)-G^*_{n-2}(t)\right) + F_{n-2}(t)G^*_{n-1}(t) \ . \end{aligned}$$

If $H_n(t) = F_n(t)G^*_{n-1}(t) - F_{n-1}(t)G^*_n(t)$ then

$$H_n(t) = -H_{n-1}(t) = \ldots = (-1)^n H_0(t) = (-1)^{n+1} I \ .$$

Now

$$G_n(t)G^*_{n-1}(t) = G_{n-1}(t)Z_n^{-1}(t)G^*_{n-1}(t) + G_{n-2}(t)G^*_{n-1}(t) \ .$$

So $G_n(t)G^*_{n-1}(t)$ is symmetric if $G_{n-1}(t)G^*_{n-2}(t)$ is symmetric. But $G_0(t)G^*_{-1}(t) = 0$ is symmetric, and so $G_n(t)G^*_{n-1}(t)$ is symmetric for all n . Therefore,

$$\begin{aligned} &F_n(t)G_n^{-1}(t) - F_{n-1}(t)G_{n-1}^{-1}(t) \\ &\qquad = \left[F_n(t)G^*_{n-1}(t)-F_{n-1}(t)G_{n-1}^{-1}(t)G_n(t)G^*_{n-1}(t)\right]G^{*-1}_{n-1}(t)G_n^{-1}(t) \\ &\qquad = \left[F_n(t)G^*_{n-1}(t)-F_{n-1}(t)G_{n-1}^{-1}(t)G_{n-1}(t)G^*_n(t)\right]G^{*-1}_{n-1}(t)G_n^{-1}(t) \\ &\qquad = (-1)^{n-1}\left(G_n(t)G^*_{n-1}(t)\right)^{-1} \ . \end{aligned}$$

LEMMA 5. *Suppose, as in Lemma* 3, *for each* n *there exists a compact interval* J_n , *with* t_0 *an interior point, and bounds* α_n, β_n *for which*

$$\left.\begin{aligned} 0 < \alpha_n^{-1}I \le \frac{\nu_n Z_n(t)}{t-t_0} \le \beta_n I \ , \quad n > 1 \ , \quad t \in J_n \ , \\ \textit{and} \quad 0 < \alpha_1^{-1}I \le \frac{\nu_1 Z_1(t)}{t-t_0} \le \frac{\beta I}{|t-t_0|} \ \text{on} \ \ J \ . \end{aligned}\right\} \tag{20}$$

Then there also exist intervals $J_{i,n}$ *, and bounds* γ_n, δ_n, h_n, k_n *for which*

$$0 < \gamma_n I \le \nu_n G_{n-1}^{-1}(t) G_n(t) \left(t-t_0\right) \le \delta_n I \ , \quad n = 2 \ , \tag{21}$$

if $t \in J_{1,n}$ *, and*

$$0 < h_n I \le \nu_n Z_1(t) G_n(T) G_{n-1}^*(t) Z_1(t) \left(t-t_0\right)^{2n-3} \le k_n I \ , \quad t \in J_{1,n}; \tag{22}$$

and

$$\begin{aligned} \left|G_n(t)\right| &= O\left(\left|t-t_0\right|^{-n}\right) \ , \\ \left|G_n^{-1}(t)\right| &= O\left(\left|t-t_0\right|^{n-1}\right) \ , \\ \left|F_n(t)\right| &= O\left(\left|t-t_0\right|^{1-n}\right) \ , \\ \left|F_n^{-1}(t)\right| &= O\left(\left|t-t_0\right|^{n-1}\right) \ , \\ \left|Z_1(t) G_n(t)\right| &= O\left(\left|t-t_0\right|^{1-n}\right) \ . \end{aligned}$$

So

$$0 < k_n^{-1} Z_1^2(t) \le (-1)^{n-1} \frac{\nu_n\left(T_n(t)-T_{n-1}(t)\right)}{\left(t-t_0\right)^{2n-3}} \le h_n^{-1} Z_1^2(t) \le \alpha^2 h_n^{-1} I$$

on $J_{1,n}$ *and*

$$0 < a_n Z_1^2(t) \le (-1)^n \frac{\nu_{n-1}\left(T_n(t)-T_{n-2}(t)\right)}{\left(t-t_0\right)^{2n-5}} \le b_n Z_1^2(t) \le \alpha^2 b_n$$

for some constants a_n, b_n .

Proof. Let $P_n(t)$ be a solution of

$$P_n(t) = P_{n-1}(t)Z_n^{-1}(t) + P_{n-2}(t)$$

with $P_2(t)P_1^*(t)$ symmetric, and $Q_n(t) = P_{n-1}^{-1}(t)P_n(t)$. Suppose on an interval $J_{1,n-1}$, $Q_{n-1}(t)$ exists and is symmetric, and there are constants $\gamma_{n-1}, \delta_{n-1} : 0 < \gamma_{n-1}I \le \nu_{n-1}Q_{n-1}(t)(t-t_0) \le \delta_{n-1}I$. Then $Q_n(t) = Z_n^{-1}(t) + Q_{n-1}^{-1}(t)$ and

$$\begin{aligned}(t-t_0)\nu_nQ_n(t) &\ge \beta_n^{-1}I - I(t-t_0)^2\gamma_{n-1}^{-1}\\ &\ge \gamma_nI\end{aligned}$$

on $J_{1},_{n} = t : |t-t_0| < \sqrt{\frac{\gamma_{n-1}}{2\beta_n}}$, $J_{1,n} \subseteq J_{1,n-1}$ and $\gamma_n = \frac{1}{2\beta_n}$.

And

$$\begin{aligned}(t-t_0)\nu_nQ_n(t) &\le \alpha_nI + I(t-t_0)^2\gamma_{n-1}^{-1}\\ &= \delta_nI\end{aligned}$$

on $J_{1,n}$, where $\delta_n = \alpha_n + \frac{1}{2\beta_n}$. Suppose $P_n(t) = G_n(t)$. Then $P_2(t)P_1^*(t) = Z_1^{-1}(t) + Z_1^{-1}(t)Z_2^{-1}(t)Z_1^{-1}(t)$ is symmetric, and on J_1 , $Q_1(t) = Z_1^{-1}(t)$, $Q_2(t) = Z_1(t) + Z_2^{-1}(t)$ so

$$0 < \gamma_2I \le (t-t_0)\nu_2Q_2(t) \le \delta_2I$$

on $J_{1,2}$ where $J_{1,2} = t : |t-t_0| \le \beta_2^{-1}\beta_1^{-1}$, $\gamma_2 = \beta_2 - \beta_1$. $\delta_2 = \alpha_2 + \beta_1$. $Q_2(t)$ exists and is symmetric; so $Q_n(t)$ exists and is symmetric and $0 < \gamma_nI \le (t-t_0)\nu_nQ_n(t) \le \delta_nI$ on $J_{1,n}$, for all $n > 2$. Then $G_n(t) = Z_1^{-1}(t)Q_2(t) \ldots Q_n(t)$ so $|G_n(t)| = O\left(|t-t_0|^{-n}\right)$ and $|Z_1(t)G_n(t)| = O\left(|t-t_0|^{1-n}\right)$. And $Z_1(t)G_n(t)G_{n-1}^*(t)Z_1(t) = Q_2(t) \ldots Q_n(t) \ldots Q_2(t)$. So $|Z_1(t)G_n(t)G_{n-1}^*(t)Z_1(t)| \le k_n|t-t_0|^{3-2n}$ on $J_{1,n}(t)$ for some constants $k_n = \delta_2^2 \ldots \delta_{n-1}^2\delta_n$. And

$$\left|\left(Z_1(t)G_n(t)G^*_{n-1}(t)Z_1(t)\right)^{-1}\right| \le h_n^{-1}\,|t-t_0|^{2n-3} .$$

On $J_{1,n}$, $\nu_n Q_n(t)\sigma(t) > 0$. So

$$0 \le h_n I \le \nu_n Z_1(t)G_n(t)G^*_{n-1}(t)Z_1(t)\left(t-t_0\right)^{2n-3} \le k_n I \qquad (22)$$

on $J_{1,n}$. Also $G_n^{-1}(t) = Q_n^{-1}(t) \dots Q_2^{-1}(t)Z_1(t)$. So

$$\begin{aligned}|G_2^{-1}(t)| &\le \gamma_n^{-1} \dots \gamma_2^{-1}\beta_1|t-t_0|^{n-1} \text{ on some interval}\\ &= O\left(|t-t_0|^{n-1}\right) .\end{aligned}$$

If $Q_n(t) = F_{n-1}^{-1}(t)F_n(t)$, then $F_n(t) = Q_2(t) \dots Q_n(t)$, since $F_1(t) = I$ and $Q_2(t) = Z_2^{-1}(t)$, so

$$0 < \gamma_2 I = \beta_2^{-1} I \le \nu_2 Q_2(t)\left(t-t_0\right) \le \alpha_2 I = \delta_2 I$$

on $J_{1,2} = J_2$, so $0 \le \gamma_n I \le \nu_n Q_n(t)\left(t-t_0\right) \le \delta_n I$ on $J_{1,n}$. So $|F_n(t)| = O\left(|t-t_0|^{1-n}\right)$, $|F_n^{-1}(t)| = O\left(|t-t_0|^{n-1}\right)$. Finally, from (22),

$$0 < k_n^{-1}Z_1^2(t) \le \nu_n\left(t-t_0\right)^{3-2n}\left(G_n(t)G^*_{n-1}(t)\right)^{-1} \le h_n^{-1}Z_1^2(t) \le \alpha h_n^{-1} I .$$

And $\left(G_n(t)G^*_{n-1}(t)\right)^{-1} = (-1)^{n-1}\left(T_n(t)-T_{n-1}(t)\right)$. Then $T_n(t) - T_{n-2}(t) = (-1)^n\left[\left(G_{n-1}(t)G^*_{n-2}(t)\right)^{-1} - \left(G_n(t)G^*_{n-1}(t)\right)^{-1}\right]$, so

$$\begin{aligned}(-1)^n\nu_n\sigma(t)\left(T_n(t)-T_{n-2}(t)\right) &\ge \left(k_{n-1}^{-1}-k_n^{-1}|t-t_0|^2\right)Z_1^2(t)|t-t_0|^{2n-3}\\ &\ge a_n Z_1^2(t)|t-t_0|^{2n-3}\end{aligned}$$

if $|t-t_0|^2 < \dfrac{k_n}{2k_n-1}$, $a_n = \dfrac{k_{n-1}}{2}$ and

$$\begin{aligned}(-1)^n\nu_n\sigma(t)\left(T_n(t)-T_{n-2}(t)\right) &\le \left(h_{n-1}^{-1}+h_n^{-1}|t-t_0|^2\right)Z_1^2(t)|t-t_0|^{2n-3}\\ &\le b_n Z_1^2(t)|t-t_0|^{2n-3}\end{aligned}$$

if $|t-t_0|^2 < \left(b_n-h_{n-1}^{-1}\right)h_n$, which completes the proof.

The convergents $T_n(t)$ are at present undefined at t_0. To make use of Riccati inequalities, it is necessary to be sure that each function $T_n(t)$ is at least continuous at t_0. Defining $T_n(t_0) = S_0$ for all $n \geq 1$, we have

LEMMA 6. *$T_n(t)$ is continuous at t_0 for all $n \geq 1$. And $R[T_n(t)] = 0$ at t_0 if $n \geq 2$.*

Proof. $T_1(t) = S_0 + \int_{t_0}^{t} A(u)du$, so $T_1(t_0)$ is continuous at t_0. If $n > 1$,

$$\begin{aligned} |T_n(t)-T_1(t)| &\leq \sum_{i=2}^{n} |T_i(t)-T_{i-1}(t)| \\ &\leq \sum_{i=2}^{n} k_i|t-t_0| \end{aligned}$$

for some constants k_i. Therefore, $T_n(t) \to S_0$ as $t \to t_0$.

$$\begin{aligned} &T_2(t) \\ &= \left(Z_1^{-1}(t)+Z_2(t)\right)^{-1} = Z_1(t)\left(I+Z_2(t)Z_1(t)\right)^{-1} \\ &= Z_1(t) - Z_1(t)Z_2(t)Z_1(t)\left(I+Z_2(t)Z_1(t)\right)^{-1} \\ &= Z_1(t) - Z_1(t)Z_2(t)Z_1(t) + Z_1(t)Z_2(t)Z_1(t)Z_2(t)Z_1(t)\left(I+Z_2(t)Z_1(t)\right)^{-1}. \end{aligned}$$

Therefore,

$$\begin{aligned} |T_2(t)-Z_1(t)+Z_1(t)Z_2(t)Z_1(t)| &= |T_2(t)-V_1(t)+V_2(t)| \\ &\leq |Z_1(t)|^3|Z_2(t)|^2\left(1-|Z_2(t)||Z_1(t)|\right)^{-1} \\ &\leq k|t-t_0|^2 \end{aligned}$$

for some constant K. Furtheremore

$$\begin{aligned} Z_1(t) = V_1(t) &= S_0 + A(t_0)(t-t_0) + \int_{t_0}^{t} \left(A(u)-A(t_0)\right)du \\ &= S_0 + A(t_0)(t-t_0) + o(|t-t_0|), \end{aligned}$$

$$Z_1(t)Z_2(t)Z_1(t) = V_2(t) = \int_{t_0}^{t} V_1(u)C(u)V_1(u)du$$

$$= S_0C(t_0)S_0(t-t_0) + o(|t-t_0|) \ .$$

Therefore,

$$\left|\frac{T_2(t)-S_0}{t-t_0} - A(t_0) + S_0C(t_0)S_0\right| \to 0$$

as $t \to t_0$, that is,

$$T_2'(t_0) = A(t_0) - S_0C(t_0)S_0 \ .$$

And

$$|T_n(t)-T_2(t)| \leq \sum_{i=3}^{n} |T_i(t)-T_{i-1}(t)|$$

$$\leq k|t-t_0|^3$$

for some constant k . Therefore,

$$\left|\frac{T_n(t)-S_0}{t-t_0} - T_2(t_0)\right| \leq \left|\frac{T_2(t)-S_0}{t-t_0} - T_2'(t_0)\right| + k(t-t_0)^2$$

$$\to 0 \quad \text{as} \quad t \to t_0 \ .$$

Therefore, $T_n'(t_0) = T_2'(t_0)$, $n \geq 2$, and

$$R[T_n(t_0)] = T_n'(t_0) - A(t_0) + T_n(t_0)C(t_0)T_n(t_0) = 0 \ .$$

The convergents as bounds to solutions

THEOREM 2. *If* $S(t)$ *is the solution of* (1) *with* $S(t_0) = S_0$ *, and* $T_n(t)$, $S(t)$ *exist on* $[t_0, t]$ *, or* $[t, t_0]$ *, then*

$$\sigma(t)\nu_{n-1}(-1)^n(S(t)-T_n(t)) \geq 0 \ . \tag{25}$$

Proof. By Lemma 6, if $n > 1$, $R[T_n(t_0)] = 0$. By Theorem 1, if $t > t_0$, $R[T_n(t)] = -K_{n,n}(t)$ and $(-1)^n\nu_{n-1}K_{n,n}(t) > 0$ if $t \neq t_0$. Therefore,

$$(-1)^n \nu_{n-1} R\left[T_n(t)\right] \le 0 \ , \quad T_n(t_0) = S(t_0) \, ,$$

and by application of Theorem 1 of Chapter 2 on inequalities,

$$\sigma(t)(-1)^n \nu_{n-1}\left(S(t) - T_n(t)\right) \ge 0 \ .$$

Finally,

$$S(t) - T_1(t) = -\int_{t_0}^{t} T_1(u)C(u)T_1(u)du$$

so $-\nu_0 \sigma(t)\left(S(t)-T_1(t)\right) \ge 0$.

The rather cumbersome definitions involved in the expression of Theorem (1) can be restated in terms of the elements of the solution of the recurrence relations (14) with a gain in simplicity and results about the relationships of the approximants $\{T_n(t)\}$ to each other.

LEMMA 7. *Where* $K_{n,n}(t)$ *is defined*

$$K_{n,n}(t) = (-1)^n\left(T_n(t)G_{n-1}(t)-F_{n-1}(t)\right)B_{n-1}(t)\left(G^*_{n-1}(t)T_n(t)-F^*_{n-1}(t)\right) \quad (26)$$

where $B_n(t)$ *has been defined in* (6) *as alternatively* $A(t)$ *and* $C(t)$.

Proof. By definition (8),

$$K_{n,n}(t) = (-1)^n M_{n,n}(t)A_n(t)M^*_{n,n}(t) \ .$$

If n is even,

$$K_{n,n}(t) = M_{n,n}(t)V_n(t)V^{-1}_{n-1}(t) \ \ldots \ V^{-1}_1(t)B_{n-1}(t)V^{-1}_1(t) \ \ldots \ V_n(t)M^*_{n,n}(t)$$

and if n is odd,

$$K_{n,n}(t) = -M_{n,n}(t)V_n(t) \ \ldots \ V_1(t)B_{n-1}(t)V_1(t) \ \ldots \ M^*_{n,n}(t) \ .$$

Let

$$J_{n,r}(t) = (-1)^r M_{n,r}(t)V_r(t)V^{-1}_{r-1}(t) \ \ldots \ V_1^{(-1)^{r-1}}(t) \ .$$

Now $M_{n,r}(t) = H_{n,1}(t) \ \ldots \ H_{n,r}(t)$ and

$$H_{n,r}(t)H_{n,r+1}(t)V_{r+1}(t) = \left(I-H_{n,r}(t)\right)V_r(t) \ . \tag{10}$$

So

$$M_{n,r+1}(t) = \left(M_{n,r-1}(t)-M_{n,r}(t)\right)V_r(t)V_{r+1}^{-1}(t)$$

and

$$J_{n,r+1}(t) = J_{n,r-1}(t) + M_{n,r}(t)V_{r-1}(t)V_{r-2}^{-1}(t) \ \dots \ V_1^{(-1)^r}(t) \ .$$

But

$$Z_r(t) = V_1^{(-1)^{r-1}} \ \dots \ V_{r-1}^{-1}(t)V_r(t)V_{r-1}^{-1}(t) \ \dots \ V_1^{(-1)^{r-1}}(t)$$

so

$$J_{n,r}(t) = M_{n,r}(t)V_{r-1}(t) \ \dots \ V_1^{(-1)^r}(t)Z_r(t)$$

and

$$J_{n,r+1}(t) = J_{n,r-1}(t) + J_{n,r}(t)Z_r^{-1}(t) \ .$$

So, in the suffix r , $J_{n,r+1}(t)$ is a solution of the second order linear difference equation (14).

But $F_r(t)$, $G_r(t)$ are also linearly independent solutions, and

$$\begin{aligned} J_{n,1}(t) &= -H_{n,1}(t)V_1(t) = -T_n(t) \\ &\qquad\qquad\qquad\quad = -T_n(t)G_0(t) + F_0(t) \ ; \end{aligned}$$

$$\begin{aligned} J_{n,2}(t) &= H_{n,1}(t)H_{n,2}(t)V_2(t)V_1^{-1}(t) \\ &= I - H_{n,1}(t) \quad \text{from (10)} \\ &= I - T_n(t)V_1^{-1}(t) \\ &= F_1(t) - T_n(t)G_1(t) \ . \end{aligned}$$

Therefore, $J_{n,r}(t) = F_{r-1}(t) - T_n(t)G_{r-1}(t)$ and $J_{n,n}(t) = F_{n-1}(t) - T_n(t)G_{n-1}(t)$. But

$$K_{n,n}(t) = J_{n,n}(t)B_{n-1}(t)J^*_{n,n}(t)(-1)^n$$

so (26) is established.

The convergents as approximants to solutions

The simplification introduced by Lemma 7 allows an estimate for the difference between a solution $S(t)$ of (1) and its bounds $T_n(t)$. The bounds are a good approximation near the point t_0 about which the expansion is made.

THEOREM 3. *Suppose on some compact interval* J *containing* t_0 *, that* $S(t)$ *exists and* $G_n(t)$ *is invertible, so* $T_n(t)$ *exists. Then there exist constants* a, b *for which*

$$0 < aZ_1^2(t) \leq (-1)^n \nu_{n-1} \frac{\left(S(t)-T_n(t)\right)}{\left(t-t_0\right)^{2n-1}} \leq bZ_1^2(t)$$

on J . *Since* $Z_1(t)$ *is bounded,* $S(t) - T_n(t) = O\left(t-t_0\right)^{2n-1}$.

Proof.

$$\begin{aligned} T_n(t)G_{n-1}(t) - F_{n-1}(t) &= G_n^{*-1}(t)\left(F_n^*(t)G_{n-1}(t)-G_n^*(t)F_{n-1}(t)\right) \\ &= (-1)^{n-1}G_n^{*-1}(t) \end{aligned}$$

from Lemma 4. So $K_{n,n}(t) = (-1)^n G_n^{*-1}(t)B_n(t)G_n^{-1}(t)$. From (1) and (9),

$$\begin{aligned} \left(S(t)-T_n(t)\right)' = {} & \tfrac{1}{2}\left(S(t)-T_n(t)\right)C(t)\left(S(t)+T_n(t)\right) \\ & + \tfrac{1}{2}\left(S(t)+T_n(t)\right)C(t)\left(S(t)-T_n(t)\right) + K_{n,n}(t) . \end{aligned}$$

Let $\theta(t)$ be a fundamental matrix of $\theta' = -\tfrac{1}{2}C(t)\left(S(t)+T_n(t)\right)\theta$. Then

$$S(t) - T_n(t) = (-1)^n \theta^{*-1}(t) \int_{t_0}^{t} \theta^*(u)G_n^{*-1}(u)B_n(u)G_n^{-1}(u)\theta(u)du\theta^{-1}(t) .$$

But $\theta(u)$, $\theta^{-1}(t)$ and $B_n(u)$ are bounded on J , by γI , say. Then

$$|S(t)-T_n(t)| \le \gamma\left|\int_{t_0}^{t} |G_n^{-1}(u)|^2 du\right| \le \gamma_1|t-t_0|^{2n-1}$$

from Lemma 5, since $|G_n^{-1}(u)| = O\left(|u-t_0|^{n-1}\right)$. But

$$\begin{aligned} S(t) - T_n(t) &= S(t) - T_{n+2}(t) + T_{n+2}(t) - T_n(t) \quad \text{near } t_0 \\ &= T_{n+2}(t) - T_n(t) + O\left(|t-t_0|^{2n+3}\right) . \end{aligned}$$

So by Lemma 5,

$$0 \le aZ_1^2(t) \le \nu_{n-1}(-1)^n \frac{\left(S(t)-T_n(t)\right)}{\left(t-t_0\right)^{2n-1}} \le bZ_1^2(t)$$

since $T_{n+2}(t) - T_n(t)$ obeys a similar inequality, and $Z_1^{-1}(t) = O\left(|t-t_0|^{-1}\right)$.

Other forms for the Riccati equations for the convergents - monotomy relations

The result of Lemma 7 can be re-expressed in a number of ways, each with a different use.

Let $D_n(t) = (-1)^n B_{n-1}(t)$, so $(-1)^n \nu_{n-1} D_n(t) > 0$. Then $T_n(t)$ is a solution of each of the following Riccati equations:

$$\begin{aligned} R_{1,n}[W] &= R[W] + \left(WG_{n-1}(t)-F_{n-1}(t)\right)D_n(t)\left(G^*_{n-1}(t)W-F^*_{n-1}(t)\right) \\ &= 0 ; \end{aligned} \tag{29}$$

$$\begin{aligned} R_{2,n}[W] &= R[W] + \left(WG_{n-2}(t)-F_{n-2}(t)\right)Z_n(t)D_n(t)Z_n(t)\left(G^*_{n-2}(t)W-F^*_{n-2}(t)\right) \\ &= 0 ; \end{aligned} \tag{30}$$

$$\begin{aligned} R_{3,n}[W] &= R[W] + \left(G_n^{-1}(t)\right)^* D_n(t)G_n^{-1}(t) \\ &= 0 . \end{aligned} \tag{31}$$

LEMMA 8. *If* $G_n(t)$ *is invertible, so* $T_n(t)$ *exists, and also* $T_{n-1}(t)$ *,* $T_{n-2}(t)$ *exist, then the following Riccati equations and inequalities hold if* $t \ne t_0$ *:*

$$R_{i,n}\left[T_n(t)\right] = 0 \text{ , } \quad i = 1 \ldots 3 \text{ ,}$$

$$R_{1,n}\left[T_{n-2}(t)\right] = 0 \text{ ,} \tag{32}$$

$$(-1)^n \nu_n R_{1,n}\left[T_{n-1}(t)\right] > 0 \text{ ,} \tag{33}$$

$$(-1)^{n-1} \nu_{n-1} R_{2,n}\left[T_{n-2}(t)\right] > 0 \text{ .} \tag{34}$$

Proof. $R_{1,n}\left[T_n(t)\right] = 0$ was established in the proof for Theorem 3. And $R_{3,n}\left[T_n(t)\right] = 0$ follows from Lemma 7 and Theorem 1. Now

$$\begin{aligned} T_{n-2}(t)G_{n-1}(t) - F_{n-1}(t) &= \left(T_{n-2}(t)G_{n-2}(t)-F_{n-2}(t)\right)Z_{n-1}^{-1}(t) \\ &\quad + T_{n-2}(t)G_{n-3}(t) - F_{n-3}(t) \quad \text{from (18)} \\ &= T_{n-2}(t)G_{n-3}(t) - F_{n-3}(t) \text{ .} \end{aligned}$$

So $R_{1,n}\left[T_{n-2}(t)\right] = R_{1,n-2}\left[T_{n-2}(t)\right] = 0$.

Now

$$G_{n-1}(t) = \left(G_n(t)-G_{n-2}(t)\right)Z_n(t) \text{ ,}$$

$$F_{n-1}(t) = \left(G_n(t)-G_{n-2}(t)\right)Z_n(t) \text{ ,}$$

from (18). So

$$\begin{aligned} T_n(t)G_{n-1}(t) - F_{n-1}(t) &= \left[T_n(t)G_n(t)-T_n(t)G_{n-2}(t)-F_n(t)+F_{n-2}(t)\right]Z_n(t) \\ &= -\left[T_n(t)G_{n-2}(t)-F_{n-2}(t)\right]Z_n(t) \text{ .} \end{aligned}$$

Therefore, $R_{2,n}\left[T_n(t)\right] = 0$.

Now $R_{1,n}\left[T_{n-1}(t)\right] = R\left[T_{n-1}(t)\right]$ since $T_{n-1}(t)G_{n-1}(t) - F_{n-1}(t) = 0$ and

$$\begin{aligned} R_{1,n-1}\left[T_{n-1}(t)\right] &= 0 \\ &= R\left[T_{n-1}(t)\right] + \left(G_{n-1}^*(t)\right)^{-1}D_{n-1}(t)G_{n-1}^{-1}(t) \text{ .} \end{aligned}$$

So $R_{1,n}\left[T_{n-1}(t)\right] = -\left(G_{n-1}^*(t)\right)^{-1}D_{n-1}(t)G_{n-1}^{-1}(t)$. Therefore, $(-1)^n \nu_n R_{1,n}\left[T_{n-1}(t)\right] > 0$. Likewise

$$\begin{aligned} R_{2,n}\left[T_{n-1}(t)\right] &= R\left[T_{n-2}(t)\right] \\ &= R_{1,n-1}\left[T_{n-2}(t)\right] \text{ .} \end{aligned}$$

Therefore, $(-1)^{n-1}\nu_{n-1}R_{2,n}\left[T_{n-2}(t)\right] > 0$.

Lemma 8 is the device by which it can be shown that the sequences $\{T_n(t)\}$ of convergents have monotonocity properties, as in the following theorem.

THEOREM 4. *If* $Z_1(t)$ *is invertible, and* $S(t)$ *exists, between* t_0 *and* t *, then* $T_n(t)$, $T_{n-1}(t)$ *and* $T_{n-2}(t)$ *exist also and*

$$\sigma(t)\nu_n(-1)^{n-1}\big(T_n(t)-T_{n-1}(t)\big) > 0 \ , \tag{35}$$

$$\sigma(t)\nu_{n-1}(-1)^{n}\big(T_n(t)-T_{n-2}(t)\big) > 0 \ . \tag{36}$$

Proof. By Lemma 5, in some neighbourhood of t_0 ,

$$\begin{aligned}(-1)^{n-1}\nu_n\sigma(t)\big(T_n(t)-T_{n-1}(t)\big) &= \big(G_n(t)G^*_{n-1}(t)\big)^{-1}\nu_n\sigma(t)\\ &> 0\end{aligned}$$

if $t \neq t_0$ and

$$\begin{aligned}&(-1)^{n}\nu_{n-1}\sigma(t)\big(T_n(t)-T_{n-2}(t)\big)\\ &\qquad = \nu_{n-1}\sigma(t)\Big[\big(G_{n-1}(t)G^*_{n-2}(t)\big)^{-1}-\big(G_n(t)G^*_{n-1}(t)\big)^{-1}\Big]\\ &\qquad > 0\end{aligned}$$

for t close to t_0 .

If $\nu_n = \nu_{n-1}$, so $A(t)$ and $C(t)$ have the same sign, then $T_n(t)$ exists everywhere, if $Z_1(t)$ is invertible. Consequently the above inequalities can be extended using the Riccati inequalities of the previous lemma.

Otherwise $S(t)$ satisfies

$$R_{1,n}[S(t)] = \big(S(t)G_{n-1}(t)-F_{n-1}(t)\big)D_n(t)\big(G^*_{n-1}(t)S(t)-F^*_{n-1}(t)\big)$$

and $(-1)^n\nu_{n-1}R_{1,n}[S(t)] \geq 0$. And $R_{1,n}\left[T_{n-1}(t)\right](-1)^n\nu_{n-1} < 0$ from (33), since $\nu_n = -\nu_{n-1}$.

Therefore, by Theorem 1 of Chapter 2, $T_n(t)$ exists as long as $S(t)$ and $T_{n-1}(t)$ exist as t moves away from t_0, provided $Z_1(t)$ remains invertible, and

$$(-1)^n \nu_{n-1}\sigma(t)\left(S(t)-T_n(t)\right) \geq 0 \ ,$$

$$(-1)^n \nu_{n-1}\sigma(t)\left(T_n(t)-T_{n-1}(t)\right) > 0 \ .$$

Likewise $T_{n-1}(t)$ exists as long as $S(t)$, $T_{n-2}(t)$ exist, and so on. But $T_1(t)$ always exists.

Therefore the requirement that $T_{n-1}(t)$ exist is unnecessary, and the theorem is proved.

Bounds and the inverse equation

If the coefficient matrices are interchanged, (1) becomes

$$U' = C(t) - UA(t)U \tag{37}$$

which is the inverse equation of (1): If $S(t)$ is a solution of (1) invertible at t, then $S^{-1}(t)$ is a solution of (37).

We denote by a bar, the derived matrices corresponding to (37); for example $\bar{V}_1(t) = U\left(t_0\right) + \int_{t_0}^{t} C(u)du$, $\bar{\nu}_1 C(u) > 0$ etc.

The previous results all apply to (37) and yield a series of bounds and approximants to $U(t)$. If $U\left(t_0\right)$ is invertible, $S_0 = U^{-1}\left(t_0\right)$ and $S(t)$ is a solution of (1) with $S\left(t_0\right) = S_0$, then the dual bounds $\bar{T}_n(t)$ are related also to $S(t)$; Theorem 3 applies directly to show that

$$\begin{aligned} \left|\bar{T}_n^{-1}(t)-S(t)\right| &= \left|\bar{T}_n^{-1}(t)\left(\bar{T}_n(t)-U(t)\right)U^{-1}(t)\right| \quad \text{where } S(t) = U^{-1}(t) \\ &\leq \left|\bar{T}_n^{-1}(t)\right| \left|\bar{T}_n(t)-U(t)\right| \left|U^{-1}(t)\right| \\ &= O\left(\left|t-t_0\right|^{2n-1}\right) \ . \end{aligned} \tag{38}$$

So if $S(t_0)$ is invertible, $S(t)$ has two alternative continued fraction expansions about t_0 .

Now, using (28) and (9),

$$\bar{T}'_n(t) = C(t) - \bar{T}_n(t)A(t)\bar{T}_n(t) - (-1)^n\left[\bar{G}_n^{-1}(t)\right]^*\bar{B}_{n-1}(t)\bar{G}_n^{-1}(t)$$

so

$$\left[\bar{T}_n^{-1}(t)\right]' = A(t) - \bar{T}_n^{-1}(t)C(t)\bar{T}_n^{-1}(t) + (-1)^n\left[\bar{F}_n^{-1}(t)\right]^*\bar{B}_{n-1}(t)\bar{F}_n^{-1}(t) \quad (39)$$

since $\bar{G}_n^{-1}(t)\bar{T}_n^{-1}(t) = \bar{F}_n^{-1}(t)$.

In the following lemma we show that the bounds $U_n(t) = \bar{T}_n^{-1}(t)$ derived from the inverse equation of (1) also have useful relationships with solutions of (1); these bounds are particularly useful because they apply to solutions of (1) which may not exist at t_0 , and particularly to maximal and minimal solutions on intervals open-ended at t_0 .

LEMMA 9. *Let* $S(t)$ *be a solution of* (1) *existing in a half-open interval* $J = (t_0, a]$ *or* $[a, t_0)$. *Suppose* $\bar{F}_n(t)$ *exists and is invertible on* J *, and let* $U_n(t) = \bar{G}_n(t)\bar{F}_n^{-1}(t)$. *Then there exist constants* a_n, b_n *for which*

$$0 < a_nI \le (-1)^n\bar{\nu}_{n-1}\frac{\left(S(t)-U_n(t)\right)}{\left(t-t_0\right)^{2n-1}} \le b_nI \quad (40)$$

and

$$0 < (-1)^n\bar{\nu}_{n-1}\sigma(t)\left(S(t)-U_n(t)\right) .$$

Proof. The proof is like that of Theorem 3. For $t \ne t_0$, let $S(t) = S(t) - U_n(t)$.

Then

$$\begin{aligned} X(t) &= S(t)\left(U(t)-\overline{T}_n(t)\right)U_n(t) \\ &= X(t)\left(U(t)-\overline{T}_n(t)\right)U_n(t) + U_n(t)\left(U(t)-\overline{T}_n(t)\right)U_n(t) \ . \end{aligned}$$

But $\left(U(t)-\overline{T}_n(t)\right)U_n(t) = O|t-t_0|$, and $U_n(t)$ is bounded near t_0 . So $X(t) = O|t-t_0|$. Therefore $X\left(t_0\right)$ can be defined to be zero. From (1) and (39),

$$\begin{aligned} X'(t) &= \left(S(t)-U_n(t)\right)' = \\ &= U_n(t)C(t)U_n(t) - S(t)C(t)S(t) - (-1)^n\overline{F}_n^{*-1}(t)\overline{B}_{n-1}(t)\overline{F}^{-1}(t) \end{aligned}$$

so

$$\begin{aligned} X'(t) = -X(t)C(t)\tfrac{1}{2}\left(U_n(t)+S(t)\right) &- \tfrac{1}{2}\left(U_n(t)+S(t)\right)C(t)X(t) \\ &- (-1)^n\overline{F}_n^{*-1}(t)\overline{B}_{n-1}(t)\overline{F}_n^{-1}(t) \ . \end{aligned}$$

Let $\theta(t)$ be a fundamental matrix of $\theta' = \tfrac{1}{2}C(t)\left(U_n(t)+S(t)\right)\theta$. Then

$$X(t) = -(-1)^n\theta^{*-1}(t)\int_{t_0}^{t} \theta^*(u)\overline{F}_n^{*-1}(u)\overline{B}_{n-1}(u)\overline{F}_n^{-1}(u)\theta(u)du\theta^{-1}(t) \ .$$

But $\theta(u)$, $\theta^{-1}(u)$, $\overline{B}_{n-1}(u)$ and $\overline{B}_{n-1}^{-1}(u)$ are bounded in a compact neighbourhood J of t_0 . Suppose γ is a bound to all of them. Then if $t \in J$,

$$\begin{aligned} |X(t)| &\le \gamma^5\left|\int_{t_0}^{t} |\overline{F}_n^{-1}(u)|^2du\right| \\ &\le b_n \, |t-t_0|^{2n-1} \end{aligned}$$

using Lemma 5, where b_n is a positive constant.

Conversely, for some vector ξ , let $\xi(u) = \overline{F}_n^{-1}(u)\theta(u)\theta^{-1}(t)\xi$. Then $|\xi| \le |\overline{F}_n(u)||\theta^{-1}(u)||\theta(t)||\xi(u)|$, so

$$|\xi(u)| \ge \gamma_2|u-t_0|^{2(n-1)}$$

by Lemma 5. And $\overline{\nu}_{n-1}\overline{B}_{n-1}(u) \geq \gamma^{-1}I$. But

$$-\sigma(t)(-1)^n\overline{\nu}_{n-1}\xi^*X(t)\xi = \sigma(t)\int_{t_0}^{t} \xi^*(u)\overline{\nu}_{n-1}\overline{B}_{n-1}(u)\xi(u)du$$

$$\geq \gamma^{-1}\sigma(t)\int_{t_0}^{t} \xi^*(u)\xi(u)du$$

$$\geq a_n|t-t_0|^{2n-1}|\xi|^2$$

where $a_n = \gamma_2^2\gamma^{-1}$. So $-(-1)^n\sigma(t)\overline{\nu}_{n-1}\dfrac{X(t)}{|t-t_0|^{2n-1}} \geq a_nI > 0$,

proving the lemma.

LEMMA 10. *On any interval including* t_0 *where* $U_n(t)$ *and* $U_{n+1}(t)$ *both exist*

$$(-1)^n\overline{\nu}_{n-1}\sigma(t)\big(U_{n+1}(t)-U_n(t)\big) > 0 . \tag{42}$$

Proof. $U_{n+1}(t) - U_n(t) = U_{n+1}(t) - S(t) + S(t) - U_n(t)$ near t_0 and $U_{n+1}(t) - S(t) = O\big(t-t_0\big)^{2n+1}$. But

$$(-1)^n\overline{\nu}_{n-1}\sigma(t)\big(S(t)-U_n(t)\big) > \tfrac{1}{2}a_nI|t-t_0|^{2n-1}$$

by Lemma 9.

So in some neighbourhood of t_0 , $\big(U_{n+1}(t)-U_n(t)\big)(-1)^n\overline{\nu}_{n-1}\sigma(t) > 0$. But

$$U_{n+1}(t) - U_n(t) = \big(\overline{F}^*_{n+1}(t)\big)^{-1}\overline{G}^*_{n+1}(t) - \overline{G}_n(t)\overline{F}_n^{-1}(t)$$

$$= \big(\overline{F}^*_{n+1}(t)\big)^{-1}\big(\overline{G}^*_{n+1}(t)\overline{F}_n(t)-\overline{F}^*_{n+1}(t)\overline{G}_n(t)\big)\overline{F}_n^{-1}(t)$$

$$= (-1)^{n-1}\big(\overline{F}_n(t)\overline{F}^*_{n+1}(t)\big)^{-1} .$$

But $\big(\overline{F}_n(u)\overline{F}^*_{n+1}(u)\big)^{-1}$ is certainly invertible on $\big(t_0, t\big]$ or $\big[t, t_0\big)$ so the sign of $U_{n+1}(t) - U_n(t)$ cannot change. Therefore, $(-1)^n\overline{\nu}_{n-1}\sigma(t)\big(U_{n+1}(t)-U_n(t)\big) > 0$ wherever $U_{n+1}(t)$, $U_n(t)$ exist.

THEOREM 5. *Let* $\hat{S}(t)$ *be the maximal solution of* (1) *at* t_0 *, so* $U(t) = S^{-1}(t)$ *,* $t \neq t_0$ *, is the solution of* (35) *with* $U(t_0) = 0$ *. Suppose* $\nu_n = (-1)^n \nu_0$ *so that* $A(t)$ *and* $C(t)$ *have opposite signs. Then* $\hat{S}(t)$ *exists on* $[t_0, t_1]$ *only if* $U_n(t)$ *exists on* $[t_0, t_1]$ *for each* n *.*

Proof. From Lemma 9, $\nu_0(\hat{S}(t)-U_n(t)) \geq 0$, since $\overline{\nu}_{n-1} = \nu_n = (-1)^n \nu_0$ provided $\hat{S}(t)$, $U_n(t)$ exist for $t_0 \leq t \leq t_1$. And from Lemma (10), $\nu_0(U_n(t)-U_1(t)) \geq 0$, $n \geq 1$, where $U_1(t) = \left(\int_{t_0}^{t} C(u)du\right)^{-1}$ exists for all $t \neq t_0$. So if $U_n(t)$ does not exist on an interval, then $\nu_0 U_n(t)$ must become unbounded in a positive direction so $\nu_0 \hat{S}(t)$ must also become infinitely large.

COROLLARY. *A test for oscillation (non-disconjugacy) for the Hamiltonian system corresponding to* (1) *is therefore to enquire whether* $\overline{F}_n$ *is non-invertible at some point in* $[t_0, t_1]$ *say. If so, then* (1) *has no solution on* $[t_0, t_1]$.

As an example, if this test is applied to the equation $y' = 1 + y^2$, with the approximants $U_n(t)$ determined about $t_0 = 0$, then the upper bound for the first right-hand conjugate point of 0 , determined by the first positive zero of $\overline{F}_n(t)$, is for

$n = 3$,	3.87 ,
$n = 4$,	3.24 ,
$n = 5$,	3.153 ,
$n = 6$,	3.1425 ,
$n = 7$,	3.14165 ,

the actual point being $\pi = 3.14159 \ldots$.

Convergence

On the matter of convergence, we have not been able to prove that under any general circumstances the sequence $\{T_n(t)\}$ of convergents of the continued fraction converge to a solution of (1). However, at least in the case where $A(t)$ and $C(t)$ are of opposite sign, $\{T_n(t)\}$ converges to some limit $T(t)$ wherever $S(t)$ exists, since it is a monotone sequence. And so

$$\left(T_n(t)-T_{n-1}(t)\right)(-1)^{n-1} = \left(G_n(t)G^*_{n-1}(t)\right)^{-1} \to 0 \quad \text{as} \quad n \to \infty \; .$$

If $G_n^{-1}(t) \to 0$ uniformly for all n , then

$$R\{T_n(t)\} = (-1)^n\left(G_n^*(t)\right)^{-1}B_n(t)G_n^{-1}(t) \to 0$$

uniformly so $R\{T(t)\} = 0$, and $T(t) = S(t)$.

So it is the gap between the knowledge that

$$\left(G_n(t)G^*_{n-1}(t)\right)^{-1} \to 0$$

and the requirement that $\left(G_n(t)\right)^{-1} \to 0$ uniformly that needs to be filled.

The autonomous equation

If $A(t)$ and $C(t)$ are positive definite and independent of t , convergence can readily be proved for continued fraction solutions. Expanding the solution which has value zero at $t = 0$, we have $Z_n(t) = B_n \frac{t}{2n-1}$, (where $B_0 = C$, $B_1 = A$, and $B_n = C$ if n is even, A if n is odd). And $G_n(t)G^*_{n-1}(t)$ is a polynomial in t with positive definite matrices as coefficients, and its leading term is

$$\left(\frac{ACA\ldots ACAt^{2n-1}}{(2n-1)(2n-3)^2\ldots 1}\right)^{-1}$$

with $(2n-1)$ matrices in the product.

So

$$|T_n(t)-T_{n-1}(t)| \le \frac{\gamma^n\gamma^{n-1}t^{2n-1}}{\lfloor\underline{2n-5}} \quad \text{if} \quad |A| = \alpha,\ |C| = \gamma\ ,$$
$$\to 0 \quad \text{as} \quad n \to \infty\ , \text{ for all } \ t\ .$$

Useful results can also be obtained if (1) is derived from an autonomous equation with linear terms as in Chapter 3. However, the situation then is more complex, and although continued fraction solutions have promising application, we shall not pursue the question here.

Relaxation of the sign requirements for the coefficients

One way of avoiding the requirement that $A(t)$, as well as $C(t)$, in (1) should be positive or negative definite, is to transform the Riccati equation. Let

$$V_1(t) = K + \int_{t_0}^{t} A(u)du \tag{43}$$

for a constant symmetric matrix K , and $U(t) = W(t) - V_1(t)$, where $W(t)$ is a solution of (1).

Then

$$\begin{aligned} U'(t) &= -W(t)C(t)W(t) \\ &= -U(t)C(t)U(t) - V_1(t)C(t)U(t) - U(t)C(t)V_1(t) \\ &\qquad\qquad - V_1(t)C(t)V_1(\ )\ . \end{aligned} \tag{44}$$

Let $M(t)$ be an invertible solution of

$$M' = C(t)V_1(t)M \tag{45}$$

and $S(t) = M^*(t)U(t)M(t)$. Then

$$S'(t) = -S(t)M^{-1}(t)C(t)M^{*^{-1}}(t)S(t) - M^*(t)V_1(t)C(t)V_1(t)M(t) \tag{46}$$

which is of the same form as (1), and the coefficients have opposite signs. $S(t)$ can then be approximated by a continued fraction.

In return for the assured existence of the coefficients needed to construct a continued fraction, a linear equation (45) has to be solved, which sacrifices the explicit character of the original solution.

However, with more complex definitions we can proceed directly even without special properties for $A(t)$. Because of the greater complexity we give only the basic definitions here. A pair of sequences $L_{2n+1}(t)$ and $V_{2n}^{-1}(t)$ can be defined, for $t \neq t_0$:

$$L_{2n+1}(t) = - L_{2n-3}(t)V_{2n-2}^{-1}(t)V_{2n}(t)$$

$$+ L_{2n-1}(t)V_{2n}^{-1}(t)\left[\begin{array}{l}\displaystyle\int_{t_0}^{t} V_{2n}(u)V_{2n-2}^{-1}(u)L_{2n-3}^{*}(u)C(u)L_{2n-1}(u)du \\ \displaystyle+ \int_{t_0}^{t} L_{2n-1}^{*}(u)C(u)L_{2n-3}(u)V_{2n-3}^{-1}(u)V_{2n}(u)du\end{array}\right] ,$$

$$V_{2n}(t) = \int_{t_0}^{t} L_{2n-1}^{*}(u)C(u)L_{2n-1}(u)du$$

and

$$L_{-1}(t)V_0^{-1}(t) = I ,$$

$$L_1(t) = V_1(t) = S_0 + \int_{t_0}^{t} A(u)du .$$

We can impose some local conditions at t_0 to ensure that $L_{2n-1}(u)\xi$ is not identically zero in a neighbourhood of t_0 for any vector $\xi \neq 0$, and to ensure that the integrands in the definition of $L_{2n+1}(t)$ are integrable near t_0 . $A(t_0) > 0$ would be an adequate condition, but it seems likely that much weaker conditions would serve.

If the sequence $\{V_n(t)\}$ as defined earlier, exists here, then $L_{2n+1}(t) = V_1(t)V_2^{-1}(t) \ldots V_{2n+1}(t)$. But otherwise, $V_{2n+1}(t)$ may not exist for $n > 0$.

We define $Z_{2n+1}(t) = L_{2n+1}(t)V_{2n}^{-1}(t)L_{2n-1}^{*}(t)$ and $Z_{2n}^{-1}(t) = L_{2n-1}(t)V_{2n}^{-1}(t)L_{2n-1}^{*}(t)$. Define $H_{2n+1}(t)$ by

$$H_{2n+1}(t) = V_2^{(-1)^{n-1}}(t)V_4^{(-1)^n}(t) \ldots V_{2n}^{-1}(t)$$

$$\times \left[V_{2n}(t) + \int_{t_0}^{t} V_{2n}(u)V_{2n-2}^{-1}(u)L_{2n-3}^{*}(u)C(u)L_{2n-1}(u)du + \int_{t_0}^{t} L_{2n-1}^{*}(u)C(u)L_{2n-3}(u)V_{2n-2}^{-1}(u)V_{2n}(u)du \right] \times$$

$$\times V_{2n}^{-1}(t) \ldots V_2^{(-1)^{n-1}}(t)$$

and $H_1(t) = V_1(t)$.

Then the approximants $T_{2n+1}(t)$ are given by

$$T_{2n+1}(t) = H_1(t) - \big(H_3 - \big(H_5 - \ldots H_{2n+1}(t)\big)^{-1}\big)^{-1} .$$

We can define

$$G_{2n+1}^{-1}(t) = L_{2n+1}(t)V_{2n}^{-1}(t)V_{2n-2}(t) \ldots$$

$$\ldots V_2^{(-1)^n}(t)\big(T_3(t)-T_1(t)\big)^{(-1)^n} \ldots \big(T_{2n+1}(t)-T_{2n-1}(t)\big) .$$

Then $R\big[T_{2n+1}(t)\big] = G_{2n+1}^{*-1}(t)C(t)G_{2n+1}^{-1}(t)$.

The proof of this last result is more difficult than in Theorem 1. However it is very useful, because it is a Riccati equation in $T_{2n+1}(t)$ which involves only terms known to exist, like $L_{2n+1}(t)$ and terms like $\big(T_3(t)-T_1(t)\big)^{(-1)^n} \ldots \big(T_{2n-1}(t)-T_{2n-3}(t)\big)^{-1}$ whose existence can be established sequentially.

The sequence $\{T_{2n+1}(t)\}$ is once again monotonic, and bounded by the solution $S(t)$ of (1) being approximated. Consequently it can be shown that $T_{2n+1}(t)$ exists between t and t_0 if $S(t)$ exists on the same interval.

A necessary condition for the existence of solutions of the Riccati equation

Let

$$V_1(u, v) = \int_u^v A(w)dw \ , \quad \overline{V}_1(u, v) = \int_u^v C(w)dw \ ,$$

$$\overline{V}_2(u, v) = \int_u^v \overline{V}_1(u, t)A(t)\overline{V}_1(u, t)dt$$

where $u, v \in [t_0, t_1]$.

THEOREM 6. *If* $C(t) > 0$ *on* $[t_0, t_1]$ *, and* (1) *has a solution existing on* (t_0, t_1) *, then if* $t_0 < u \le v < t_1$ *,*

$$\overline{V}_1^{-1}(t_0, u) + \overline{V}_1^{-1}(t_0, u)\overline{V}_2(t_0, u)\overline{V}_1^{-1}(t_0, u) + V_1(u, v)$$
$$\ge \overline{V}_1^{-1}(t_1, v) + \overline{V}_1^{-1}(t, v)\overline{V}_2(t_1, v)\overline{V}_1^{-1}(t_1, v) \ . \qquad (46)$$

If $A(t) < 0$ in $[t_0, u], [v, t_1]$ then $\overline{V}_2(t_0, u) < 0$, $\overline{V}_2(t_1, v) > 0$ and Theorem 6 reduces to Theorem 5 of Chapter 3.

Proof. If any solution of (1) exists on (t_0, t_1) there is a maximal solution $S_+(t)$ and a minimal solution $S_-(t)$. Near t_0 , $S_+^{-1}(t)$ is the solution of $V' = C(t) - VA(t)V$ with $V(t_0) = 0$.

Let $\overline{V}_1^{-1}(u, v)(\overline{V}_1(u, v)+\overline{V}_2(u, v))\overline{V}_1^{-1}(u, v) = U_2(u, v)$.

In some interval $[t_0, a]$, $0 < \alpha I \le C(u) \le \beta I$.

Then $0 < \alpha(t-t_0)I \le \overline{V}_1(t_0, t) \le \beta(t-t_0)I$. Therefore, $\overline{V}_2(t_0, t) = O(|t-t_0|^3)$. Let $\overline{T}_2(t_0, t) = U_2^{-1}(t_0, t)$. $\overline{T}_2$ is well defined for t close to t_0 , since in some neighbourhood of t_0 ,

$$\overline{V}_1(t_0, t) \ge \alpha(t-t_0)I > -\overline{V}_2(t_0, t)$$

and so $\overline{V}_1(t_0, t) + \overline{V}_2(t_0, t)$ is invertible. Then

$$\overline{T}_2(t_0, t) = \overline{V}_1(t_0, t)\left(I-\left(\overline{V}_1(t_0, t)+\overline{V}_2(t_0, t)\right)^{-1}\overline{V}_2(t_0, t)\right)$$
$$= \overline{V}_1(t_0, t) + O\left(|t-t_0|^3\right) .$$

Therefore, $\overline{T}_2(t_0, t_0) = 0$ and

$$\frac{d}{dt}\overline{T}_2(t_0, t)_{t=t_0} = C(t_0) = C(t_0) - \overline{T}_2(t_0, t_0)A(t_0)\overline{T}_2(t_0, t_0) .$$

Elsewhere,

$$\overline{T}_2'(t_0, t) = C(t) - \overline{T}_2(t_0, t)A(t)T_2(t_0, t)$$
$$- \left[\overline{T}_2(t_0, t)\overline{V}_1^{-1}(t_0, t) - I\right]C(t)\left[\overline{V}_1^{-1}(t_0, t)\overline{T}_2(t_0, t) - I\right]$$

and $S_+^{-1\prime}(t) = C(t) - S_+^{-1}(t)A(t)S_+^{-1}(t)$. So $S_+^{-1}(t) \geq \overline{T}_2(t_0, t) > 0$ near t_0 . Now

$$U_2'(t_0, t) = A(t) - U_2C(t)U_2 + \left[\overline{V}_1^{-1}(t_0, t)-U_2\right]C(t)\left[\overline{V}_1^{-1}(t_0, t)-U_2\right]$$

and near t_0 , $S_+(t) \leq \overline{T}_2^{-1}(t_0, t) = U_2(t_0, t)$. Therefore, $S_+(t) \leq U_2(t_0, t)$ wherever $S_+(t)$ exists. And

$$S_+(v) - S_+(u) = V_1(u, v) - \int_u^v S_+(t)C(t)S_+(t)dt$$

so $S_+(v) \leq V_1(u, v) + S_+(u)$ on $[u, v]$ if $S_+(t)$ exists on $[u, v]$. By an argument like the earlier one in this proof, $S_-(v) \geq U_2(t_1, v)$. But $S_-(v) \leq S_+(v)$, that is, $U_2(t_1, v) \leq V_1(u, v) + U_2(t_0, u)$. QED

Further tests for disconjugacy can be devised on an individual basis.

Notes for Chapter 4

As this thesis was in the final stages of preparation for submission, a paper appeared by W.G. Fair [*1*] called "Continued fraction solutions to the Riccati equation". It deals with Riccati equations in a Banach algebra, and generalises a paper by E.P. Merkes and W.T. Scott [*1*]. It is unfortunate that we have not had sufficient time to investigate in detail the relationship between this chapter, and Fair's approach, which is from a quite different direction. Fair deals with the unsymmetric Riccati equation with analytic coefficients, and his expansions are derived from the coefficients of the power series expansions of the matrix coefficients. They do not have any significance as bounds, and are treated formally, and may not necessarily approximate solutions.

Fair concludes with some remarks on significance for control theory, in the autonomous case.

A form of the bound $\overline{T}_2(t)$ appears in a proof of a stability theorem of Kalman ([*2*], p. 114-115). This proof contains some errors, the most difficult to rectify being in the second part, which could have been treated as a dual of the first. A different version, making more explicit use of the bounds $T_2(t)$ and $\overline{T}_2(t)$ was given by W.A. Coppel in a seminar series at the Australian National University. A similar application is given in Corollary 2 to Theorem 2 of our next chapter.

It is interesting that in the linear regulator problem posed in a simple form in the next chapter, $\overline{T}_2(t)$ and its function as a bound can be derived by application of a constant (non-optimal) control, and comparison of the value of the criterion function with its optimal value.

Bucy [*1*] produces a bound which is, in effect of the type $V_1 + (\overline{V}_1)^{-1}$ (Lemma 4). His proof is however, defective, since it uses the assumption that if A, B, C are symmetric, $0 \leq A \leq B$ and $C > 0$, then $ACA \leq BCB$. This is false even when $C = I$. However, a bound of the type $\overline{T}_2^{-1}(t)$ can be used to establish his later theorems.

For some recent papers using a quasilinearization approach to finding bounds for the solutions of the autonomous matrix Riccati equation see Bellman [*1*], Aoki [*1*], Kleinman [*1*] and McClamroch [*1*].

In connection with a test for oscillation due to Tomastik [*1*], Barrett [3, p. 504] raises the question of whether, it was necessary to impose sign requirements on both coefficients. Theorem 5 of Chapter 3 allows a partial relaxation, and appears to go as far as is possible with criteria of that kind. The test in Theorem 6 above only uses the sign of one coefficient $C(t)$.

Index of definitions in Chapter 4

CHAPTER 5

ASYMPTOTIC BEHAVIOUR OF RICCATI EQUATIONS AND THEIR ASSOCIATED LINEAR SYSTEMS

Introduction

This chapter investigates the behaviour of symmetric solutions of the Riccati equation

$$W' = A(t) - WC(t)W \tag{1}$$

where $A(t)$, $C(t)$ are symmetric and non-negative definite on $(-\infty, \infty)$ and of solutions of the associated linear equation

$$y' = C(t)W(t)y \tag{2}$$

where $W(t)$ is a symmetric solution of (1), and is particularly concerned with the asymptotic behaviour of solutions of (2).

We assume everywhere

$$\text{controllability} \quad [C] : \int_a^b C(t)dt > 0 \quad \text{if} \quad a > b \text{ , and}$$

$$\text{observability} \quad : \int_a^b A(t)dt > 0 \quad \text{if} \quad a > b \text{ .}$$

These conditions could be relaxed, particularly by assuming $\int_a^b A(t)dt > 0$ if $b - a$ exceeds some minimum value, but this would be a diversion from our chief purpose.

Any solution $W(t)$ of (1) with $W\left(t_0\right) \geq 0$ exists in $\left[t_0, \infty\right)$ since

$$0 \leq W(t) \leq W\left(t_0\right) + \int_{t_0}^t A(u)du \text{ .}$$

Since $[C]$ holds, there is a minimal solution $\hat{W}(t)$ of (1) on $\left[t_0, \infty\right)$, using a dual version of Theorem 1, Chapter 3.

Although we assume less about $A(t)$ and $C(t)$ than in the previous chapter, properties of the bounds $T_1(t)$, $T_2(t)$, $T_3(t)$,

$\overline{T}_1(t)$, $\overline{T}_2(t)$ and $\overline{T}_3(t)$ are unaffected. In particular, if $\Pi(t, t_0)$ is the solution of (1) with $\Pi(t_0, t_0) = 0$, then

$$\begin{aligned}\hat{W}(t_1) &\le \Pi(t_1, t_1+\varepsilon)\\ &\le V_1(t_1)\left(V_1(t_1)+V_2(t_1)\right)^{-1}V_1(t_1)\end{aligned}$$

where

$$V_1(t) = -\int_t^{t_1+\varepsilon} A(u)du < 0$$

if $\varepsilon > 0$, and

$$V_2(t) = -\int_t^{t_1+\varepsilon} V_1(u)C(u)V_1(u)du \le 0 \ ;$$

so $\hat{W}(t_1) < 0$.

If $W(t)$ is a solution of (1) and $y(t)$ a solution of (2), it is fairly easy to get information about the behaviour of $y^*(t)W(t)y(t)$ which acts as a Lyapunov function for (2). This is shown in Lemma 1, which is directly applied in Theorems 1 and 2. The behaviour of solutions $W(t)$ of (1) is, in a sense, like that of the coefficients $A(t)$ and $C(t)$; if $A(t)$ and $C(t)$ are constant, then $W(t)$ tends to a constant non-zero limit as $t \to \infty$, and if $A(t)$, $C(t)$ are polynomial functions, then at worst eigenvalues of $W(t)$ tend to zero or infinity asymptotically with a finite pwoer of t . Furthermore the behaviour of $W(t)$ is usually influenced mostly by the nearby values of $A(t)$ and $C(t)$. This situation contrasts with that of the linear equation (2), whose solutions behave in a way most unlike that of the coefficient functions. And if $n = 1$,

$$y(t) = y(t_0)\, e^{\int_{t_0}^{t} C(u)W(u)du} ,$$

so asymptotic behaviour for linear equations is affected as much by distant values of the coefficients as by nearby values.

These remarks are not made in a spirit of rigor, but to motivate our strategy for determining the asymptotic behaviour of (2). Having found inequalities for $y^*(t)W(t)y(t)$, it is then necessary to find bounds (possibly time-varying) for $W(t)$. This is done in, for example, the corollaries to Theorem 2 below.

Information about the asymptotic behaviour of (2) is often important, as in control theory applications, for example. But it also has implications for the behaviour of solutions of (1) relative to each other; Theorem 3 shows that the set of solutions of (1), excluding those which differ from the principal solution by a singular matrix, have a tendency to aggregate as $t \to \infty$, and the rapidity of this aggregation is determined by the behaviour of corresponding solutions of (2).

To summarise the applicable parts of the results of this chapter, Theorems 1 and 2 give information about solutions $y(t)$ of (2) near infinitely, provided that the corresponding solution of (1), $W(t)$, is bounded in some way. The bounds of the previous chapter give information about the behaviour of $W(t)$; Corollary 2 to Theorem 2 shows how the bounds can be used in a certain situation with particular application to the case of a uniformly controllable and observable system. No simple general criteria for the eventual asymptotic behaviour of solutions of (1) are given, but an example at the end of this chapter illustrates a method of approach using the results of the previous chapter.

The linear regulator problem

We present a rather simplified version of the linear regulator problem; it nevertheless has the important features of the usual, more general version, to which our arguments also apply.

Problem. Let $y(t)$ be a "state" vector, $z(t)$ a "control" vector, with $y'(t) = C(t)z(t)$. Find a "control law" or function $k(x, t)$ for which $z(t) = k\big(y(t), y\big)$ minimises the function

$$V_z\big(y_0, t_0, t_1\big) = y^*(t)Py(t) + \int_{t_0}^{t_1} \big(y^*(u)A(u)y(u)+z^*(u)C(u)z(u)\big)du \ .$$

$A(t)$ and $C(t)$ are assumed symmetric and positive definite on $\left[t_0, t_1\right]$, $y\left(t_0\right) = y_0$, and P_1 is also assumed symmetric.

Solution. $k(x, t) = W(t)x$, so $z(t) = W(t)y(t)$, where $W(t)$ is the solution of

$$W' = A(t) - WC(t)W \tag{1}$$

with $W(t_1) = P_1$. Then $V(y_0, t_0, t_1) = y^*(t_0)W(t_0)y(t_0)$.

A solution certainly exists for $t_0 < t_1$ if $P_1 \le 0$, so $W(t_1) \le 0$.

On an infinite interval (with $t_1 = \infty$), the function to be minimised is

$$\int_{t_0}^{\infty} \left(y^*(u)A(u)y(u)+z^*(u)C(u)z(u)\right)du \; .$$

The solution is $k(x, t) = \overline{W}(t)x$, where

$$\overline{W}(t) = \lim_{t_1\to\infty} W(t, t_1) \; ,$$

where $W(t, t_1)$ is the solution of (1) with $W(t_1, t_1) = 0$.

Then

$$V(y_0, t_0, t_1) = y^*(t_0)\overline{W}(t_0)y(t_0) \; .$$

$\overline{W}(t)$ can be shown to exist and be negative definite for each t . In fact, $\overline{W}(t)$ is the inverse of the minimal solution (on $(-\infty, \infty)$) of the inverse equation $V' = C(t) - VA(t)V$. We also show that under certain conditions, there is only one negative definite solution of (1) on $(-\infty, \infty)$, so $\hat{W}(t)$ is also the minimal solution of (1).

For solutions of the infinite interval problem, it is necessary to know whether solutions of the equation

$$y' = C(t)\overline{W}(t)y \tag{2}$$

are stable, and also whether $\overline{W}(t)$ is a stable solution of the Riccati equation (1).

We show that solutions of (2) are, under fairly general conditions, stable (as a consequence of the negativity of $\overline{W}(t)$), but that $\overline{W}(t)$ is not a stable solution of (1). This conclusion was reached by Kalman [2] in the case where the coefficients are uniformly controllable and observable, that is, behave like functions bounded away from 0 and ∞ .

Asymptotic behaviour of the linear equation

The following lemma is basic.

LEMMA 1. *If* $P(t)$ *is a solution of* (1) *existing on* $[b, c]$, $\Pi(t)$ *a function of* t *differentiable on* $[b, c]$, $y(t)$ *a solution on* $[b, c]$ *of*

$$y' = C(t)P(t)y \tag{2}$$

and

$$Q(t) = [P(t)-\Pi(t)]C(t)[P(t)-\Pi(t)] - R[\Pi(t)] \tag{3}$$

then

$$\begin{aligned} y^*(c)P(c)y(c) &= y^*(b)P(b)y(b) + y^*(c)\Pi(c)y(c) - y^*(b)\Pi(b)y(b) \\ &\qquad + \int_b^c y^*(t)Q(t)y(t)dt \qquad (4) \\ &\geq y^*(b)P(b)y(b) + y^*(c)\Pi(c)y(c) - y^*(b)\Pi(b)y(b) \\ &\qquad - \int_b^c y^*(t)R[\Pi(t)]y(t)dt \ . \qquad (5) \end{aligned}$$

Proof.

$$\begin{aligned} &\frac{d}{dt}[y^*(t)\big(P(t)-\Pi(t)\big)y(t)] \\ &\quad = y^*(t)\left[\begin{array}{c} P(t)C(t)\big(P(t)-\Pi(t)\big)+\big(P(t)-\Pi(t)\big)C(t)P(t) \\ +A(t)-P(t)C(t)P(t)-\Pi'(t) \end{array}\right]y(t) \\ &\quad = y^*(t)\left[\begin{array}{c} \big(P(t)-\Pi(t)\big)C(t)\big(P(t)-\Pi(t)\big)-\Pi(t)C(t)\Pi(t)+A(t) \\ -\Pi'(t) \end{array}\right]y(t) \\ &\quad = y^*(t)Q(t)y(t) \ . \end{aligned}$$

The stated result follows on integration.

COROLLARY. *If* $\Pi(t)$ *is a solution of* (1) *with* $\Pi(b) = 0$ *, then* $Q(t) \geq 0$ *, so*

$$y^*(c)P(c)y(c) \geq y^*(b)P(b)y(b) + y^*(c)\Pi(c)y(c) \ . \tag{6}$$

The next theorem gives information about the asymptotic behaviour of $y^*(t)P(t)y(t)$. In investigating stability of (2), this acts as a Lyapunov function.

THEOREM 1. *If* $P(t)$ *is a solution of* (1), $y(t)$ *a corresponding solution of* (2) *on* $[b, c]$ *and if* $m(t)$ *is a continuous function throughout* $[b, c]$ *for which*

$$m(t)P(t) \leq A(t) + P(t)C(t)P(t) \tag{7}$$

then

$$y^*(c)P(c)y(c) \geq y^*(b)P(b)y(b)e^{\int_b^c m(u)du} . \tag{8}$$

Remark. If $P(b) > 0$, $m(t) > 0$, then (8) says $|y^*(c)P(c)y(c)|$ is an increasing function of c . If $P(b) < 0$, $P(c) < 0$, then $|y^*(c)P(c)y(c)|$ is decreasing as c increases.

Proof. Let $\Pi(t) \equiv 0$ in Lemma 1, and so

$$Q(t) = A(t) + P(t)C(t)P(t) . \tag{9}$$

Then

$$y^*(c)P(c)y(c) = y^*(b)P(b)y(b) + \int_b^c y^*(t)Q(t)y(t)dt . \tag{10}$$

Let

$$s(t) = y^*(b)P(b)y(b) + \int_b^t y^*(u)Q(u)y(u)du . \tag{11}$$

Then

$$\begin{aligned} s'(t) &= y^*(t)Q(t)y(t) \\ &\geq m(t)y^*(b)P(t)y(t) \\ &= m(t)s(t) \end{aligned}$$

by (10) and (11). Therefore

$$\frac{d}{dt}\left\{s(t)e^{-\int_b^t m(u)du}\right\} \geq 0$$

and so $s(t) \geq s(b)e^{\int_b^t m(u)du}$ and $s(b) = y^*(b)P(b)y(b)$. Therefore,

$$y^*(t)P(t)y(t) = s(t) \geq y^*(b)P(b)y(b)e^{\int_b^t m(u)du} .$$

COROLLARY 1. (7) holds if either $m(t)P(t) \leq A(t)$ or $m(t)P^{-1}(t) \leq C(t)$. Alternatively if $m_1(t)P(t) \leq A(t)$, and $P(t)$ is invertible on $[b, c]$, and $m_2(t)P^{-1}(t) \leq C(t)$, then (7) and (8) hold with $m(t) = m_1(t) + m_2(t)$.

COROLLARY 2. If $p(t)I \geq P(t) \geq q(t)I$, then

$$p(c)|y(c)|^2 \geq q(b)|y(b)|^2e^{\int_b^c m(u)du} .$$

If $p(t) > 0$, and $A(t) \geq a(t)I \geq 0$ we can take $m(t) = \frac{a(t)}{p(t)}$. If $q(t) < 0$, we can take $m(t) = \frac{a(t)}{q(t)}$.

So the asymptotic behaviour of $y^*(t)P(t)y(t)$ can be expressed in terms of a relation between $P(t)$ and the coefficients $A(t)$ and $C(t)$. However the behaviour of $P(t)$, as a solution of (1), should be more closely related to the integrals of the coefficient matrices over some intervals, and the following theorem is introduced with this in mind. Its proof is, to some extent, a discrete analogue of the proof of Theorem 1.

Let $\Pi(u, v)$ be a function with $\Pi(v, v) = 0$,

$$\frac{d\Pi}{du}(u, v) = A(u) - \Pi(u, v)C(u)\Pi(u, v) .$$

THEOREM 2. *Suppose* $P(t)$ *is a solution of* (1) *existing on* $[t_0, \infty)$ *and that* $\{t_n\}, \{d_n\}$ *are sequences of real numbers for which*

$$0 < d_nP(t_n) \leq \Pi(t_n, t_{n-1}) ,$$

$d_n < 1$ *and* $t_n > t_{n-1}$ *for all* $n = 1, 2, \ldots, p$, $t_1 = b$, $t_p = c$. *Then*

$$y^*(c)P(c)y(c) \geq y^*(b)P(b)y(b) \prod_{i=2}^{p} \frac{1}{1-d_i} . \tag{13}$$

Proof. By Corollary 1 to Lemma 1,

$$\begin{aligned} y^*(t_n)P(t_n)y(t_n) &\geq y^*(t_{n-1})P(t_{n-1})y(t_{n-1}) + y^*(t_n)\Pi(t_n, t_{n-1})y(t_n) \\ &\geq y^*(t_{n-1})P(t_{n-1})y(t_{n-1}) + d_n y^*(t_n)P(t_n)y(t_n) . \end{aligned}$$

Let $c_n = y^*(t_n)P(t_n)y(t_n)$. Then $(1-d_n)c_n \geq c_{n-1}$. So

$$c_p \geq \prod_{i=2}^{p} \frac{1}{1-d_i} c_1 ;$$

that is,

$$y^*(c)P(c)y(c) \geq y^*(b)P(b)y(b) \prod_{i=2}^{p} \frac{1}{1-d_i} .$$

COROLLARY 1. If

$$\frac{P(t_n)}{a_n} \leq I \leq \frac{\Pi(t_n, t_{n-1})}{b_n}$$

for sequences $\{a_n\}$, $\{b_n\}$ then we can take $d_n = \frac{b_n}{a_n}$ in (13). In this case, if $P(t)$ is a positive solution,

$$a_n|y(t_n)|^2 \geq y^*(t_n)P(t_n)y(t_n) \geq y^*(t_1)P(t_1)y(t_1) \prod_{i=2}^{p} \frac{a_i}{a_i-b_i} .$$

If $0 < a_n \leq a$, $b_n \geq b > 0$, we can take

$$d_n = \frac{b}{a} , \tag{14}$$

and then $y^*(t_i)P(t_i)y(t_i) \to \infty$ exponentially as $i \to \infty$. If $d_n \geq \alpha > 0$ then again

$$y^*(t_p)P(t_p)y(t_p) \geq y^*(t_1)P(t_1)y(t_1) \frac{1}{(1-\alpha)^{p-1}}$$

$$\to \infty \text{ exponentially as } p \to \infty .$$

But if $d_n \leq -\alpha < 0$, so $P(t_n) < 0$, $n = 1, 2, \ldots$ then

$$|y^*(t_p)P(t_p)y(t_p)| \le |y^*(t_1)P(t_1)y(t_1)| \frac{1}{(1+\alpha)^{p-1}} \tag{15}$$

$$\to 0 \quad \text{exponentially as} \quad p \to \infty \ .$$

COROLLARY 2. If there are positive constants α and β for which

$$\alpha I \le \int_{t_{n-1}}^{t_n} C(u)du \le \beta I \ ,$$

$$\alpha I \le \int_{t_{n-1}}^{t_n} A(u)du \le \beta I \ , \tag{16}$$

for all $n = 1, 2, \ldots$ and a monotone increasing sequence $\{t_n\}$ then

$$\Pi(t_n,\ t_{n-1}) \ge V_1(t_n)\left(V_1(t_n)+V_2(t_n)\right)^{-1}V_1(t_n)$$

from Theorem 2, Chapter 4, where

$$V_1(t) = \int_{t_{n-1}}^{t} A(u)du \ , \quad V_2(t) = \int_{t_{n-1}}^{t} V_1(u)C(u)V_1(u)du$$

so

$$\begin{aligned}\Pi^{-1}(t_n,\ t_{n-1}) &\le V_1^{-1}(t_n) + V_1^{-1}(t_n)V_2(t_n)V_1^{-1}(t_n) \\ &\le \alpha^{-1}I + \alpha^{-1}V_2(t_n) \ .\end{aligned}$$

But

$$\begin{aligned}V_2(t_n) &\le I \int_{t_{n-1}}^{t_n} |V_1(u)|^2|C(u)|du \\ &\le nI\beta^2 \left|\int_{t_{n-1}}^{t_n} C(u)du\right| \le n\beta^3 I\end{aligned}$$

because

$$\begin{aligned}|C(u)| &= \text{maximum eigenvalue of } C(u) \\ &\le \text{trace } C(u) \ .\end{aligned}$$

So

$$\int_{t_{n-1}}^{t_n} |C(u)|du \le \int_{t_{n-1}}^{t_n} \operatorname{tr} C(u)du = \operatorname{tr} \int_{t_{n-1}}^{t_n} C(u)du$$

$$\le n \left\| \int_{t_{n-1}}^{t_n} C(u)du \right| .$$

Therefore, $\Pi(t_n, t_{n-1}) \ge \gamma I$ where $\gamma^{-1} = \alpha^{-1} + n\alpha^{-2}\beta^3$. Let $P^{-1}(t)$ be a positive solution of $W' - C(t) + WA(t)W = 0$ on $[t_0, t_m]$. If $\psi(t, t_{n-1})$ is the solution of this equation with $\psi(t_{n-1}, t_{n-1}) = 0$, then $P^{-1}(t_{n-1}) > 0$, so $P^{-1}(t_n) > \psi(t_n, t_{n-1})$. And by the same process as for the original equation,

$$\psi(t_n, t_{n-1}) \ge \gamma I .$$

Therefore, $P^{-1}(t_n) > \gamma I$, and $P(t_n) < \gamma^{-1}I$. Therefore, $\gamma^{-1} > \gamma$ and $\gamma < 1$.

Using Corollary 1, with $a = \gamma^{-1}$, $b = \gamma$ in (14),

$$|y(t_m)|^2 \ge \left(\frac{\gamma^{-1}}{\gamma^{-1}-\gamma}\right)^{m-1} \frac{\gamma}{\gamma^{-1}} |y(t_1)|^2$$

$$= \frac{\gamma^2}{(1-\gamma^2)^{m-1}} |y(t_1)|^2 . \tag{17}$$

If $P(t)$ is a negative definite solution on $|t_1, t_{m+1}|$ then an analogous argument gives $0 < \frac{P(t_i)}{-\gamma^2} \le \Pi(t_i, t_{i-1})$ and $-\gamma^{-1}I \le P(t_i) \le -\gamma I < 0$, $i = 1 \dots m$. Then again from Theorem 2,

$$y^*(t_m)P(t_m)y(t_m) \ge y^*(t_1)P(t_1)y(t_1) \prod_{i=2}^{m} \frac{\gamma^2}{1+\gamma^2} ;$$

that is,

$$\gamma|y(t_m)| \le |y(t_1)| \frac{1}{(1+\gamma^{-2})^{m-1/2}} .$$

These results can be applied in a variety of situations, showing that as $t \to \infty$, $y(t) \to 0$ if $P(t) < 0$, $|y(t)| \to \infty$ if $P(t) > 0$. We give a specific application where $A(t)$, $C(t)$ are like constant matrices.

If the inequalities of the hypothesis of Corollary 2 hold for any pair of equally spaced points separated by a distance δ, then (1) is said to be uniformly controllable and observable. If this condition holds for some interval length δ, it holds for all larger interval lengths. In particular,

$$\alpha I \le \int_{t_0}^{t_1} C(u)du \le 2\beta I ,$$

$$\alpha I \le \int_{t_0}^{t_1} A(u)du \le 2\beta I ,$$

whenever $t_0 + \delta \le t_1 \le t_0 + 2\delta$.

So if t and s are two points for which $t+n\delta \le s \le t+(n+1)\delta$, $n \ge 1$, then from (17), if $P(t) > 0$ on $[t, s]$, and $[t, s]$ is divided into n equal intervals,

$$\begin{aligned} |y(s)| &\ge |y(t)| \frac{\gamma}{\left(1-\gamma^2\right)^{n/2}} && \text{where } \gamma^{-1} = \alpha^{-1} + 8n\alpha^{-2}\beta^3 \\ &\ge |y(t)| e^{(s-t)\theta}\gamma\sqrt{1-\gamma^2} && \text{where } \theta = \frac{1}{2\delta}\log\left(1-\gamma^2\right) \\ & && \qquad\quad \le \frac{n+1}{2(s-t)}\log\left(1-\gamma^2\right) . \end{aligned}$$

Therefore, $y(s) \to \infty$ exponentially as $s \to \infty$; $y(t) \to 0$ as $t \to -\infty$. If $P(t) < 0$ on $[t, s]$, then

$$|y(s)| \le |y(t)| \frac{1}{\sqrt{\gamma^2+1}} e^{-(s-t)\theta} \quad \text{where} \quad \theta = \frac{1}{2\delta}\log\left(1+\gamma^2\right) .$$

So $|y(s)| \to 0$ exponentially as $t \to \infty$, and in this case $y' = C(t)P(t)y$ is uniformly asymptotically stable at ∞.

Remark. The conditions of uniform controllability and observability, used by Kalman [2] and Bucy [1] in generalising the asymptotic properties of systems with constant coefficients, have no special status in our approach. In particular it is the ratio of values of $A(t)$ and $C(t)$ rather than their absolute values, which tends to influence asymptotic behaviour of the Riccati equation.

There is an example at the end of this chapter where particular application of our theorems is possible although there is no uniform controllability or observability.

Aggregation of solutions of the Riccati equation

There is a relationship between solutions of (2) and the differences of solutions of (1) which has been exploited in Chapter 3. If $Y(t)$ is a fundamental matrix of (2), $P(t)$ being a solution of (1) existing on $[b, \infty)$, and if $Q(t)$ is another solution of (1) existing on $[b, \infty)$ then

$$\frac{d}{dt} Y^*(t)\big(P(t)-Q(t)\big)Y(t) = Y^*(t)\big(P(t)-Q(t)\big)C(t)\big(P(t)-Q(t)\big)Y(t)$$

so if $U(t) = Y^*(t)\big(P(t)-Q(t)\big)Y(t)$ then

$$U'(t) = U(t)Y^{-1}(t)C(t)Y^{*-1}(t)U(t) \ . \tag{19}$$

If $U(b)$ is invertible, then, $V(t) = U^{-1}(t)$ is a solution of

$$V'(t) = -Y^{-1}(t)C(t)Y^{*-1}(t)$$

which exists on $[b, \infty)$, so $U(t)$ is invertible on $[b, \infty)$, and if $S(t) = \int_b^t Y^{-1}(u)C(u)Y^{*-1}(u)du$, then $S(t) > 0$ if $t > b$, from $[C]$ and

$$\begin{aligned} U(t) &= \big(U^{-1}(b)-S(t)\big)^{-1} \\ &= U(b)\big(S^{-1}(t)-U(b)\big)^{-1}S^{-1}(t) \ . \end{aligned} \tag{20}$$

Now $S^{-1}(t)$ is decreasing, non-negative and so has a limit K as $t \to \infty$. If $K - U(b)$ is invertible, then

$$U(t) \to U(b)\big(K-U(b)\big)^{-1} \quad \text{as} \quad t \to \infty \ . \tag{21}$$

Then $U(t)$ is bounded on $[b, \infty)$, so there exist scalars h, k , for which $hI \le U(t) \le kI$ on $[b, \infty)$. Then

$$h\big(Y(t)Y^*(t)\big)^{-1} \le P(t) - Q(t) \le k\big(Y(t)Y^*(t)\big)^{-1} \ . \tag{22}$$

Both Theorems 1 and 2 give information about the asymptotic behaviour of $(Y(t)Y^*(t))^{-1}$ as t increases. From (22) and Theorem 1 we deduce:

THEOREM 3. *If* $P(t)$ *is a positive solution of* (1) *on* $[b, \infty)$, $Y(t)$ *is an invertible solutions of* $Y' = C(t)P(t)Y$,

$K = \lim_{t\to\infty} \int_b^t Y^{-1}(u)C(u)Y^{*^{-1}}(u)du$, *(the limit must exist), and* $Q(t)$ *is a solution of* (1) *existing on* $[b, \infty)$ *for which* $P(b) - Q(b)$ *is invertible, and* $P(b) - Q(b) - Y^{*^{-1}}(b)KY^{-1}(b)$ *is invertible, and if* $m(t)$ *is a continuous function:* $m(t)P(t) \le A(t) + P(t)C(t)P(t)$ *then*

$$|P(t)-Q(t)| \le g|P(t)|e^{-\int_b^t m(u)du} \quad \textit{for some constant } g . \tag{23}$$

Proof. From (8),

$$Y^*(t)P(t)Y(t) \ge Y^*(b)P(b)Y(b)e^{\int_b^t m(u)du}$$

and we assume $Y^*(t)P(t)Y(t) \ge pI > 0$ for a constant p , since $P(b) > 0$. Then

$$P(t)e^{-\int_b^t m(u)du} \ge p(Y(t)Y^*(t))^{-1}$$

and

$$\begin{aligned}|P(t)|e^{-\int_b^t m(u)du} &\ge p|(Y(t)Y^*(t))^{-1}| \\ &\ge \frac{1}{g}|P(t)-Q(t)|\end{aligned}$$

from (22) where $g = \frac{1}{p}\max(|b|, |k|)$. QED

Remark. If $|P(t)|e^{-\int_b^t m(u)du} \to 0$, as in many important cases it does, then Theorem 3 asserts that "almost all" solutions approach $P(t)$ asymptotically as $t \to \infty$. The requirement that $P(b) - Q(b)$ is invertible is inessential, and only made for convenience; if it is not obtained, then P and Q can be compared with a third solution. The requirement that $P(b) - Q(b) - Y^{*^{-1}}(b)KY^{-1}(b)$ is invertible is, however, essential; if $Q(b) = \hat{W}(b)$, $\hat{W}$ being the principal solution at ∞, then $P(b) - Q(b) - Y^{*-1}(b)KY^{-1}(b) = 0$ from the proof of Theorem 1 of Chapter 3.

So Theorem 3 says that in this case all solutions except the principal solution and those which differ at b from the principal solution by a simgular matrix, converge in a single bundle as $t \to \infty$. The bundle contains positive solutions; the exclusion of the principal solution is not surprising, since it is the negative solution.

As an example, $y' = 1 - y^2$ has the general solution $y = \tanh(t+a)$ or $\coth(t+a)$, with $+1$ and -1 as special solutions. As $t \to \infty$,

$$\left(1 - \tanh(t+a)\right) = \frac{e^{-t-a}}{\cosh(t+a)} \to 0$$

and

$$1 - \coth(t+a) = \frac{e^{-t-a}}{\sinh(t+a)} \to 0 \; .$$

So all solutions except $y = -1$ aggregate about $+1$ as $t \to \infty$. This follows from Theorem 3 with $P(t) = 1$, $m(t) = 2$.

Results summarised for a special case, and an example

Finally to show the use of the results of this chapter, we first summarise their effect where the Riccati equation is uniformly controllable and observable, and then give an example of qualitatively similar behaviour of solutions of an equation which is not uniformly controllable or observable.

If equations (16) apply whenever $t_n - t_{n-1} = d$ for some fixed d, which is the definition of uniform controllability and observability, and $P(t)$ is a solution of (1) positive at $t = t_0$, then there is a constant g for which $P(t) > gI$, $P(t) < g^{-1}I$ if $t > t_0 + d$, from (16).

There is a positive constant h for which $|y(t)| = O\left(e^{ht}\right)$ as

$t \to \infty$, where $y(t)$ is a nontrivial solution of $y' = C(t)P(t)y$.

The principal solution $W(t)$ of (1) on $[t_0, \infty)$ is negative definite for all t. If $Q(t)$ is another solution with $Q(t_0) > W(t_0)$, then $|P(t)-Q(t)| = O(e^{2ht})$ as $t \to \infty$, where $P(t)$ is any non-negative definite solution of (1).

If $y(t)$ is a solution of $y' = C(t)W(t)y$ then $|y(t)| = O(e^{-ht})$ as $t \to \infty$, for some constant $h > 0$.

The aggregation of solution ensures that the Riccati equation with constant coefficients has only one constant positive and one negative solution at most.

Example of the application of Theorem 1 to the modified Bessel equation of order zero:

$$t^2x'' + tx' - t^2x = 0 \ , \quad t \geq 1 \ , \tag{24}$$

its equivalent Hamiltonian system

$$\begin{aligned} y' &= \frac{1}{t} z \ , \\ z' &= ty \ , \end{aligned} \tag{25}$$

and Riccati equation

$$w' = A(t) - w^2C(t) \ , \quad A(t) = t \ , \quad C(t) = \frac{1}{t} \ . \tag{26}$$

If $w(t)$ is a solution of (26) with $w(t) > 0$ in $(t-1, t)$, then $w(t) \geq T_2(t)$, where $T_2(t) = \left[V_1^{-1}(t)+V_1^{-2}(t)V_2(t)\right]^{-1}$, $V_1(u) = \int_{t-1}^{u} A(s)ds$ and $V_2(t) = \int_{t-1}^{t} V_1^2(u)C(u)du$. By elementary calculation, it can be verified that

$$\begin{aligned} T_2(t) &\geq \frac{3}{4}\,(t-\tfrac{1}{2}) \\ &\geq \tfrac{1}{2}t \end{aligned}$$

if $t \geq 2$.

On the other hand $v(t) = w^{-1}(t)$ is a solution of $v' = \frac{1}{t} - tv^2$. Operating in the same way with this dual equation, and omitting the very similar details, $v(t) \geq (2t)^{-1}$, so $w(t) \leq 2t$, $t \geq 2$. Therefore, $\frac{1}{2}w(t) \leq t = A(t)$, $\frac{1}{2}w^{-1}(t) \leq \frac{1}{t}$, so $\frac{1}{2}w(t) \leq w^2(t)C(t)$. Therefore, $w(t) \leq A(t) + w^2(t)C(t)$. Therefore, in Theorem 1, $m(t) = 1$, and

$$\begin{aligned} y^2(t)w(t) &\geq y^2(t_0)w(t_0)e^{\int_{t_0}^{t} du} \\ &= y^2(t_0)w(t_0)e^{(t-t_0)} \end{aligned}$$

where $y(t)$ is a solution of

$$y' = \frac{1}{t}w(t)y \,. \tag{27}$$

But $w(t) \leq 2t$, so $y^2(t) \geq \frac{1}{2t} e^t K(t_0)$ where $K(t_0) = y^2(t_0)w(t_0)$ if $w(t) \geq 0$ on $[t_0-1, t]$.

In the same sort of way it can be verified that for $t \geq 2$, if $w(t) < 0$ on $[t-1, t]$, then $-2t \leq w(t) \leq -\frac{t}{2}$.

Then $-w(t) \leq A(t) + w^2(t)C(t)$ and from Theorem 1, $y^2(t)w(t) \geq y^2(t_0)w(t_0)e^{-(t-t_0)}$ or, if $w(t) < 0$, $w(t_0) < 0$, then $y^2(t)|w(t)| \leq y^2(t_0)|w(t_0)|e^{-(t-t_0)}$. Therefore, $y^2(t) \leq \frac{1}{2t}e^{-t} K(t_0)$ where $K(t_0) = y^2(t_0)|w(t_0)|$.

Using Theorem 3, we conclude that the solutions of (26) aggregate about a positive solution, with one exception, and for two of the aggregating solutions $w_1(t)$ and $w_2(t)$,

$$|w_1(t)-w_2(t)| \le Kte^{-\frac{3}{2}t} \quad \text{as} \quad t \to \infty .$$

These conclusion can be verified by knowledge of the general solution of (26):

$$w(t) = \frac{\alpha tI_1(t)-\beta tK_1(t)}{\alpha I_0(t)+\beta K_0(t)}$$

where α, β are arbitrary constants, and I_0, I_1, K_0, K_1 modified Bessel functions. The minimal solution has $\alpha = 0$, and solutions of the corresponding linear equation (27) are multiples of $K_0(t)$. Other solutions of (27) for other solutions $w(t)$ of (26) are $\alpha I_0(t) + \beta K_0(t)$, $\alpha \ne 0$, which tend to ∞. As $t \to \infty$,

$$I_0(t) \sim \frac{1}{\sqrt{2\pi t}}e^t , \quad tI_1(t) \sim \sqrt{\frac{t}{2\pi}}e^t ,$$
$$K_0(t) \sim \sqrt{\frac{\pi}{2t}}e^{-t} , \quad tK_1(t) \sim \sqrt{\frac{\pi t}{2}}e^{-t} ,$$

and so the predicted behaviour of solutions can be verified.

Notes

On the role of the Riccati equation in the linear regulator problem, the stability of its solutions and the solutions that it generates, see Kalman [2]. Our approach to the stability of solutions of the Riccati equation differs from that of Kalman, and is like that of Bucy [1]. However we have not needed to invoke the Lyapunov stability theorem.

The results on uniformly controllable and observable systems that are corollaries of our Theorems 2 and 3 are given by Kalman [2, 6.10 and 7.2] and Bucy [1]. The results in this case, about aggregation of solutions are due to Bucy [1, Theorem 4], although his statement of the matter is incomplete, and in fact incorrect, since it does not observe the distinct behaviour of the principal solution. As Kalman's paper makes clear, solutions of linear equations associated with principal solutions are of great importance, being stable. Also, as observed in the notes to Chapter 4, the *a priori* bounds to solutions of the Riccati equation are incorrect, as used by Bucy in [1].

With respect to Bucy's final remarks, $\lim_{t_0 \to -\infty} \Pi(t, t_0)$ always exists (if $C(t) \geq 0$ and is controllable, $A(t) \leq 0$); this follows from our proof of the existence of a maximal solution on $(-\infty, \infty)$ which does not use variational arguments. Reid [15] has generalised the monotone properties of Bucy [1, Section 3] but has to assume the coefficient matrices absolutely continuous.

The properties of stabilizability and detectability, used by Wonham [1], [2], and Lukes [1], do not readily generalise to the non-autonomous case. Equation (1) of this chapter is derived from that of Chapter 3 by a congruence transformation; stabilizability exploits the possibility of first making a translation operation to give the resulting equation more desirable properties.

We draw attention to some very recent remarks of Fair [1], with the promise of a forthcoming paper, on approximating solutions of matrix quadratic equations by continued fractions. Our continued fraction expansions of Chapter 4 are helpful here. If the value at t_0 of the negative principal solution at $+\infty$ is wanted, and uniform controllability and observability, say, apply, then because of the aggregation of solutions it suffices to approximate the solution $\Pi(t, t_0+d)$ at t_0 , where d is positive and sufficiently large. $\Pi(t, t_0+d)$ is the solution of (1) which takes the value zero at $t_0 + d$, and can be approximated by a continued fraction expansion. This procedure avoids problems of instability of principal solutions.

If, for example, $C(t) = A(t) = 1$, the principal solution as $+\infty$ is -1 . If one tries to find its value at $t = 0$ by expanding the continued fraction of $\Pi(0, 5)$ one gets the sequence of approximations:

$$-5, -0.54, -1.21, -0.057, -1.0074, -0.99893, -1.00001, -0.99990 .$$

Another approach to finding values of the principal solution involving Newton-type approximations (quasi linearization), is given in the papers of Bellman [1], Aoki [1] and McClamrock [1].

BIBLIOGRAPHY

The references cited below are included for various reasons, which are indicated by the labels following them. The labels have the following significance:

[R] concerning the Riccati equation;

[D] concerning disconjugacy;

[O] concerning optimal control theory and controllability;

[I] concerning differential inequalities;

[G] for general references.

AHLBRANDT, C.D.

[*1*] "Disconjugacy criteria for self-adjoint differential systems", *J. Differential Equations* 6 (1969), 271-295. [D, R]

[*2*] "Equivalent boundary value problems for self-adjoint differential systems", *J. Differential Equations* 9 (1971), 420-435. [D, R]

[*3*] "Solutions of self-adjoint differential systems", *Rocky Mountain J. Math.* 2 (1972), 169-182. [D]

ANDERSON, B.D.O.

[*1*] "The testing for optimality of linear systems", *Int. J. Control* 4 (1966), 29-40. [O]

[*2*] "A system theory criterion for positive real matrices", *SIAM J. Control* 5 (1967), 171-182. [O]

ANDERSON, B.D.O. and MOORE, J.B.

[*1*] *Linear optimal control* (Prentice-Hall, Englewood Cliffs, New Jersey, 1971). [O]

AOKI, M.

[*1*] "Note on aggregation and bounds for the solution of the matrix Riccati equation", *J. Math. Anal. Appl.* 21 (1968), 377-383. [R]

ATHANS, M, FALB, P.

[*1*] *Optimal control* (McGraw-Hill, New York, 1966). [O]

ATKINSON, F.V.

[*1*] *Discrete and continuous boundary value problems* (Academic Press, New York, London, 1964). [MR31#416]. [G]

BARNETT, S.

[*1*] *Matrices in control theory* (Van Nostrand, London, 1971). [R, O]

BARRETT, J.H.

[*1*] "Matrix systems of second order differential equations", *Portugal. Math.* 14 (1955), 79-89. [D]

[*2*] "A Prüfer transformation for matrix differential equations", *Proc. Amer. Math. Soc.* 8 (1957), 510-518. [D]

[*3*] "Oscillation theory of linear differential equations", *Adv. Math.* 3 (1969), 415-509. [MR41#2113]. [D]

BELLMAN, R.

[*1*] *Introduction to matrix analysis*, 2nd ed., (McGraw-Hill, New York, 1970).

[*2*] "Upper and lower bounds for the solutions of the matrix Riccati equation", *J. Math. Anal. Appl.* 17 (1967), 373-379. [R]

BIRKHOFF, G.D. and HESTENES, M.R.

[*1*] "Natural isoperimetric conditions in the calculus of variations", *Duke Math. J.* 1 (1935), 198-286. [D]

BLISS, G.A.

[*1*] *Lectures on the calculus of variations* (University of Chicago Press, Chicago, 1946). [D]

BÔCHER, M.

[1] "Application of a method of d'Alembert to the proof of Sturm's theorems of comparison", *Trans. Amer. Math. Soc.* 1 (1960), 414-420. [D, R].

[2] "Non-oscillatory linear differential equations of the second order", *Bull. Amer. Math. Soc.* 7 (1900-01), 333-340. [D, R]

BROCKETT, R.W.

[1] *Finite-dimensional linear systems* (Wiley, New York, 1970). [O]

BUCY, R.S.

[1] "Global theory of the Riccati equation", *J. Comput. System Sci.* 1 (1967), 349-361. [O, R]

CHANDRA, J. and FLEISCHMANN, B.A.

[1] "Bounds and maximal solutions of nonlinear functional equations", *Bull. Amer. Math. Soc.* 74 (1968), 512-516. [I]

CHIELLINI, A.

[1] "Sui sistemi di Riccati", *Rend. Sem. Sci. Univ. Cagliari* 18 (1948), 44-58. [R]

COFFMAN, C.V.

[1] "Nonlinear differential equations on cones in Banach spaces", *Pacific J. Math.* 14 (1964), 9-16. [I]

COLES, W.J.

[1] "Linear and Riccati systems", *Duke Math. J.* 22 (1965), 333-338. [R].

COPPEL, W.A.

[1] "Comparison theorems for canonical systems of differential equations", *J. Math. Anal. Appl.* 12 (1965), 306-315. [D]

[2] *Stability and asymptotic behavior of differential equations* (D.C. Heath, Boston, 1965). (MR41#568). [I]

[3] *Disconjugacy* (Springer-Verlag, Berlin, 1971). [I, R, D]

[4] *Linear systems* (Lecture Notes, ANU). [R, O]

EDMUNDS, D.E.

[1] "Differential inequalities", *Proc. London Math. Soc.* 15 (1965), 361-372. [I]

ETGEN, G.J.

[1] "Oscillatory properties of certain nonlinear matrix differential systems of second order", *Trans. Amer. Math. Soc.* 122 (1966), 289-310. [D]

[2] "A note on trigonometric matrices", *Proc. Amer. Math. Soc.* 17 (1966), 1226-1232. [D]

[3] "On the determinants of solutions of second order matrix differential systems", *J. Math. Anal. Appl.* 18 (1967), 585-598. [D]

FAIR, W.G.

[1] "Continued fraction solution to the Riccati equation", *J. Math. Anal. Appl.* 39 (1972), 318-323. [R]

FORD, D.A. and JOHNSTON, C.D.

[1] "Invariant subspaces and the controllability and observability of linear dynamical systems", *SIAM J. Control* 6 (1968), 553-558. [O]

FRIEDLAND, B.

[1] "On solutions of the Riccati equation in optimization problems", *IEEE Trans. Aut. Control* AC-12 (1967), 303-304. [R, O]

GANTMACHER, F.R. and KREIN, M.G.

[1] *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme* (Springer-Verlag, Berlin, 1960). [D]

GELFAND, I.M. and FOMIN, S.V.

[1] *Calculus of variations,* (Prentice-Hall, New Jersey, 1963). [D]

HARTMAN, P.

[1] "Self-adjoint, non-oscillatory systems of ordinary, second order, linear differential equations", *Duke Math. J.* 24 (1957), 25-36. [MR18, 576]. [D]

[2] *Ordinary differential equations* (John Wiley and Sons, 1964). [D]

HARTMAN, P., and WINTNER, A.

[1] "On disconjugate differential systems", *Canad. J. Math.* 8 (1956), 72-81. [D]

HEINZ, E.

[1] "Halbbeschränktheit gewöhnlicher Differentialoperatoren höherer Ordnung", *Math. Ann.* 135 (1958), 1-49. [D]

HESTENES, M.R.

[1] "Applications of the theory of quadratic forms in Hilbert space to the calculus of variations", *Pacific J. Math.* 1 (1951), 525-581. [MR18, 759]. [D]

[2] *Calculus of variations and optimal control theory* (John Wiley and Sons, New York, 1966). [O]

HILLE, E.

[1] "Non-oscillation theorems", *Trans. Amer. Math. Soc.* 64 (1948), 234-252. [MR10, 376]. [D]

HINTON, D.B.

[1] "Disconjugate properties of a system of differential equations", *J. Differential Equations* 2 (1966), 420-437. [MR34#7856]. [D]

HOWARD, H.C.

[1] "Oscillation criteria for matrix differential equations", *Canad. J. Math.* 19 (1967), 184-199. [D]

INCE, E.C.

[1] *Ordinary differential equations* (Longmans, London, 1927). [G]

JACOBSON, D.H.

[1] "New conditions of boundedness of the solution of a matrix Riccati differential equation", *J. Differential Equations* 8 (1970), 258-263.

[2] "A general sufficiency theorem for the second variation", *J. Math. Anal. Appl.* 34 (1971), 578-589. [D, R]

JAKUBOVIC, V.A.

[1] "Oscillation properties of solutions of canonical equations", [Translation: *Amer. Math. Soc. Transl.* 42 (1964), 247-288]. [D]

[2] "Solution of an algebraic problem encountered in control theory", *Soviet Math. Dokl.* 11 (1970), 882-886. [O]

[3] "Factorization of symmetric matrix polynomials", *Soviet Math. Dokl.* 11 (1970), 1261-1264. [O]

KALMAN, R.E.

[1] "A new approach to linear filtering and prediction problems", *J. Basic Engr. (Trans. ASME)* 82D (1960), 35-45. [O, R]

[2] "Contributions to the theory of optimal control", *Bol. Soc. Mat. Mexicana* 5 (1960), 102-119. [O, R]

[3] "On the general theory of control systems", *Proc. First IFAC Congress, Moscow* (Butterworths, London, 1961). [O, R]

[4] "Mathematical description of linear dynamical systems", *SIAM J. Control* 1 (1963), 152-192. [O]

[5] "Lectures on controllability and observability", *Centro Internazionale Matematico Estivo, Sasso Marconi (Bologna)*, 1-9 July 1968 (Ediuioui Cremonese, Rome, 1969). [O]

KALMAN, R.E. and BUCY, R.S.

[*1*] "New results in linear filtering and prediction theory", *J. Basic Engr.* (*Trans. ASME*) 83D (1961), 95-108. [0, R]

KALMAN, R.E., HO, Y.C. and NARENDRA, K.S.

[*1*] "Controllability of linear dynamical systems", *Contributions to Differential Equations* 1 (1963), 189-213. [0]

KAMKE, E.

[*1*] "Zur Theorie der Systeme gewöhnlicher Differentialgleichungen, II", *Acta Math.* 58 (1932), 57-85. [I]

[*2*] "A new proof of Sturm's comparison theorems", *Amer. Math. Monthly* 46 (1939), 417-421. [D]

KLEINMAN, D.L.

[*1*] "On an iterative technique for Riccati equation computations", *IEEE Trans. Aut. Control.* AC-13 (1968), 114-115. [0, R]

KRASNOSELSKII, M.A.

[*1*] *Positive solutions of operator equations* (Noordhoff, 1964). [I]

KREINDLER, E.

[*1*] "Sensitivity of time-varying, linear optimal control systems", *J.O.T.A.* 3 (1969), 98-106. [0]

KREINDLER, E. and SARACHIK, P.E.

[*1*] "On the concepts of controllability and observability of linear systems", *IEEE Trans. Automatic Control* 9 (1964), 129-136. [0]

LAKSHMIKANTHAM, V. and LEELA, S.

[*1*] *Differential and integral inequalities* (Academic Press, 1969). [I]

LEE, E.B. and MARKUS, L.

[1] *Foundations of optimal control theory* (Wiley, New York, 1967). [O]

LEIGHTON, W.

[1] "Principal quadratic functionals", *Trans. Amer. Math. Soc.* 67 (1949), 253-274. [MR11, 603]. [D]

LEVIN, J.J.

[1] "On the matrix Riccati equation", *Proc. Amer. Math. Soc.* 10 (1959), 519-524. [R]

LIDSKII, V.B.

[1] "Oscillation theorems for canonical systems of differential equations", (Russian), *Dokl. Akad. Nauk SSSR* 102 (1955), 877-880. [MR17, 483]. [D]

LUENBERGER,

[1] *Optimization by vector space methods* (Wiley, New York, 1969). [G, I]

LUKES, D.L.

[1] "Stabilizability and optimal control", *Funkcial. Ekvac.* 11 (1968), 39-50. [O]

[2] "Optimal regulation of non-linear dynamical systems", *SIAM J. Control* 7 (1969), 75-100. [O]

LUKES, D.L. and RUSSELL, D.L.

[1] "The quadratic criterion for distributed systems", *SIAM J. Control* 7 (1969), 101-121. [O]

MACFARLANE, A.G.J.

[1] "An eigenvector solution of the optimal linear regulator problem", *J. Electron. Control* 14 (1963), 643-654. [O]

MAN, F.T.

[*1*] "The Davidson method of solution of the algebraic matrix Riccati equation", *Int. J. Control* 10 (1969), 713-719. [0, R]

MARTENSSON, K.

[*1*] "On the matrix Riccati equation", *Information Sciences* 3 (1971), 17-49. [R]

McCLAMROCH, N.H.

[*1*] "Duality and bounds for the matrix Riccati equation", *J. Math. Anal. Appl.* 25 (1969), 622-627. [0, R]

MERKES, E.P. and SCOTT, W.T.

[*1*] "Continued fraction solutions to the Riccati equation", *J. Math. Anal. Appl.* 4 (1962), 309. [R]

MLAK, W.

[*1*] "Note on maximal solutions of differential equations", *Contributions to Differential Equations* 1 (1963), 461-465. [I]

MLAK, W. and OLECH, C.

[*1*] "Integration of infinite systems of differential equations", *Ann. Polon. Math.* 13 (1963), 105-112. [I]

MOLINARI, B.P.

[*1*] "Algebraic solution of matrix linear equations in control theory", *Proc. IEE,* 116 (1969), 1748-1754. [0, R]

[*2*] "Redundancy in the optimal linear regulator", *IEEE Trans.* AC-16 (1971), 83-85. [0, R]

[*3*] "The stabilizing solution of the matrix quadratic equation", (to appear). [0]

MOORE, J.B.

[*1*] "Application of Riccati equations in systems engineering", *Inst. Engineers, Austral. Elec. Eng. Trans. EE* 5 (1969), 29-54. [0, R]

MOORE, J.B. and ANDERSON, B.D.O.

[1] "Extensions of quadratic minimisation theory", *Internat. J. Control* 7 (1968), 465-472 and 473-480. [O]

MORSE, M.

[1] "A generalization of the Sturm separation and comparison theorems in n-space", *Math. Ann.* 103 (1930), 52-69. [D]

[2] *The calculus of variations in the large* (Amer. Math. Soc. Colloq. Publ. 18. Amer. Math. Soc., Providence, Rhode Island, 1934). [D]

MORSE, M. and LEIGHTON, W.

[1] "Singular quadratic functionals", *Trans. Amer. Math. Soc.* 40 (1936), 252-286. [D]

NAGUMO, S.

[1] "Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen", *Proc. Phys.-Math. Soc. Japan* 24 (1942), 551-559. [I]

PERRON, O.

[1] "Ein neuer Existengbeweis für die Integrale der Differentialgleichung $y' = f(x, y)$ ", *Math. Ann.* 76 (1915), 471-484. [I]

PORTER, W.A.

[1] "On the matrix Riccati equation", *IEEE Trans. Aut. Control* AC-12 (1967), 746-749. [O, R]

POTTER, J.E.

[1] "Matrix quadratic solutions", *SIAM J. Appl. Math.* 14 (1966), 496-501. [O, R]

PRÜFER, H.

[1] "Neue Herleitung der Sturm-Liouvilleschen Reihenentwicklung stetiger Funktionen", *Math. Ann.* 95 (1926), 499-518. [D]

RADON, J.

[1] "Zum Problem von Lagrange", *Abh. Math. Sem. Univ. Hamburg* 6 (1928), 273-299. [D, R]

REDHEFFER,

[1] "On solutions of Riccati's equation as functions of initial values", *J. Rat. Mech. Anal.* 5 (1956), 835-848. [R]

[2] "The Riccati equation: initial values and inequalities", *Math. Ann.* 133 (1957), 235-250. [R]

[3] "Inequalities for a matrix Riccati equation", *J. Math. Mech.* 8 (1959), 349-367. [R]

REID, W.T.

[1] "A matrix differential equation of Riccati type", *Amer. J. Math.* 68 (1946), 237-246. "Addendum", 70 (1948), 250. [R]

[2] "Oscillation criteria for linear differential systems with complex coefficients", *Pacific J. Math.* 6 (1956), 733-751. [D, R]

[3] "A comparison theorem for self-adjoint differential equations of second order", *Ann. Math.* (2) 65 (1957), 197-202. [D]

[4] "Principal solutions of non-oscillatory self-adjoint linear differential systems", *Pacific J. Math.* 8 (1958), 147-169. [D]

[5] "Solutions of a Riccati matrix differential equation as functions of initial values", *J. Math. Mech.* 8 (1959), 221-230. [D, R]

[6] "Generalized linear differential systems", *J. Math. Mech.* 8 (1959), 705-726. [D]

[7] "Properties of solutions of a Riccati matrix differential equation", *J. Math. Mech.* 9 (1960), 749-770. [D, R]

[8] "Oscillation criteria for self-adjoint differential equations", *Trans. Amer. Math. Soc.* 101 (1961), 91-106. [D, R]

[*9*] "Riccati matrix differential equations and non-oscillation criteria for associated linear differential systems", *Pacific J. Math.* 13 (1963), 665-685. [D, R]

[*10*] "Principal solutions of non-oscillatory linear differential systems", *J. Math. Anal. Appl.* 9 (1964), 397-423. [D, R]

[*11*] "A matrix equation related to a non-oscillation criterion and Lyapunov stability", *Quart. Appl. Math.* 23 (1965), 83-87. [D, R]

[*12*] "A class of monotone Riccati matrix differential operators", *Duke Math. J.* 32 (1965), 689-696. [D, R]

[*13*] "Generalised linear differential systems and related Riccati matrix integral equations", *Illinois J. Math.* 10 (1966), 701-727. [D, R]

[*14*] *Ordinary differential equations* (Wiley, New York, 1971). [D, R, G]

[*15*] "Monotoneity properties of solutions of Hermitian Riccati matrix differential equations", *SIAM J. Math. Anal.* 1 (1970), 195-213. [D, R, O]

ROTH, W.E.

[*1*] "On the matrix equation "$X^2+AX+XB+C = 0$ ", *Proc. Amer. Math. Soc.* 1 (1950), 586-589. [R]

RUPP, R.D.

[*1*] "A method for solving a quadratic optimal control problem", *JOTA* 9 (1972), 238-251. [O]

SANDOR, S.

[*1*] "Sur l'équation différentielle matricielle de type Riccati", *Bull. Math. Soc. Sci. Math. Phys. R.P. Roumaine (NS)* 3 (51) (1959), 220-249. [MR23#A1863]. [R]

SCHUMITZKY, A.

[*1*] "On the equivalence between matrix Riccati equations and Fredholm resolvents", *J. Compat. System. Sci.* 2 (1968), 76-87. [R]

SILVERMAN, L.M.

[1] "Synthesis of impulse response matrices by internally stable and passive realizations", *IEEE Trans. Circuit Theory* 15 (1968), 238-245. [O]

SILVERMAN, L.M. and ANDERSON, B.D.O.

[1] "Controllability, observability and stability of linear systems", *SIAM J. Control* 6 (1968), 121-130. [O]

SILVERMAN, L.M. and MEADOWS, H.E.

[1] "Controllability and observability in time-variable linear systems", *SIAM J. Control* 5 (1967), 64-73. [O]

STERNBERG, R.L.

[1] "Variational methods and non-oscillation theorems for systems of differential equations", *Duke Math. J.* 19 (1952), 311-322. [D]

[2] "A theorem on hermitian solutions for related matrix differential and integral equations", *Portugal. Math.* 12 (1953), 135-139. [D]

SWANSON, C.A.

[1] *Comparison and oscillation theory of linear differential equations* (Academic Press, New York, 1968). [D]

SZARSKI, J.

[1] "Differential inequalities", *PWN*, (Polish Sci. Publ. Warsaw, 1965). [I]

TOMASTIK, E.C.

[1] "Singular quadratic functionals of n dependent variables", *Trans. Amer. Math. Soc.* 124 (1966), 60-76. [D, R]

[2] "Oscillation of non-linear matrix *DE*'s of second order", *Proc. Amer. Math. Soc.* 19 (1968), 1427-1431. [D, R]

[3] "Oscillation of systems of second order differential equations", *J. Differential Equations* 9 (1971), 436-442. [D, R]

VAUGHAN, D.R.

[*1*] "A negative exponential solution for the matrix Riccati equation", *IEEE Trans. Aut. Control* AC-14 (1969), 72-75. [R]

WALTER, W.

[*1*] *Differential und Integral Ungleichungen* (Springer, Berlin, 1964). [I]

WHYBURN, W.M.

[*1*] "Matrix differential equations", *Amer. J. Math.* 56 (1934), 587-592. [D]

WONHAM, W.M.

[*1*] "On pole assignment in multi-input controllable linear systems", *IEEE Trans. Aut. Control* AC-12 (1967), 660-665. [O, R]

[*2*] "On a matrix Riccati equation of stochastic control", *SIAM J. Control* 6 (1968), 681-697. [O, R]

YORKE, J.A.

[*1*] "Invariance for ordinary differential equations", *Math. Systems Theory* 1 (1967), 353-372. [I]