Stability and inertia

Biswa Nath Datta

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA

Received 27 February 1999; accepted 12 September 1999

Submitted by R.A. Brualdi

Dedicated to Hans Schneider for his outstanding and stable contributions to Stability and Inertia Theory

Abstract

The purpose of this paper is to present a brief overview of matrix stability and inertia theory. A few applications of inertia and stability theorems, and a nonspectral implicit matrix equation method for determining stability and inertia of a nonhermitian matrix are also presented. Inter-relationships between different theorems are explicitly stated, whenever appropriate. The paper concludes with some problems for future research in this area. © 1999 Elsevier Science Inc. All rights reserved.

AMS classification: 15A23; 93D05

Keywords: Inertia; Stability

1. Introduction

It is well-known that the system of differential equations

\[ \dot{x}(t) = Ax(t) \]

(1.1)

is asymptotically stable (that is, \( x(t) \to 0 \) as \( t \to \infty \)) if and only if all the eigenvalues of \( A \) have negative real parts. Similarly, the system of difference equations

\[ x(k + 1) = Ax(k), \quad k = 0, 1, \ldots, \]

(1.2)

is asymptotically stable if and only if all the eigenvalues of \( A \) have modulii less than 1.

E-mail address: dattab@math.niu.edu (Biswa Nath Datta)

© 1999 Elsevier Science Inc. All rights reserved.
In 1892, the Russian mathematician Alexandr Michailovich Lyapunov, in his doctoral dissertation "The General Problem of Stability of Motion", published a remarkable historical result on stability of nonlinear systems of differential equations which, in the case of the linear system (1.1), may be formulated in terms of the positive definite solution matrix $X$ of the matrix equation $XA + A^*X = -M$.

In 1952, Stein published a counterpart of Lyapunov’s result relating the stability of (1.2) to the equation $X - A^*XA = -M$. These equations (as well as their duals) are known as the Lyapunov and Stein equations, respectively. A brief life-history of Lyapunov appears in [107].

In many engineering applications, it may not be enough to determine if the system is stable. One often needs to monitor the rate at which the solution decays and needs to study other various transient responses of the system. The transient responses of the system (1.1) or (1.2) are governed by the region where the eigenvalues of $A$ are located in the complex plane. For example, if $\lambda = x + iy$ is an eigenvalue of $A$ and it is desired that the system (1.1) has the minimum decay rate $\alpha$, then $x < -\alpha < 0$; that is, the real part of each of the eigenvalues should be less than $\alpha$. Similarly, to deal with the oscillating behavior of the damped system of second-order differential equations

$$M\ddot{x}(t) + D\dot{x}(t) + Kx(t) = 0,$$

it is important to know what is the minimum value of the damping ratio $\rho$ and what is the minimum value of the undamped natural frequency $\omega$, where $\rho$ and $\omega$ are defined by the eigenvalues of the quadratic pencil

$$P(\lambda) = \lambda^2M + \lambda D + K$$

associated with (1.3). A typical eigenvalue $\lambda$ of the pencil is of the form $\lambda = x + iy = -\rho\omega \pm i\omega\sqrt{1 - \rho^2}$. If it is desired that the system have the minimal damping ratio $\rho$, then the eigenvalues should lie in a sector with a slope less than $(1/\rho^2 - 1)^{1/2}$. Similarly, if it is desired that the system have the minimal frequency $\omega_0$, then the eigenvalues should lie outside a circle of radius $\omega_0$. For details of these results and further engineering applications of the distribution of the eigenvalues of a matrix in a complex plane, see [61,62].

The above considerations have led to the development of several generalized stability theorems which deal with stability of a matrix $A$ with respect to regions in the complex plane that are more general than the half planes and the unit circle. At the same time, the Lyapunov and Stein stability theorems have been generalized to inertia theorems. Stability and inertia theory for operators, matrix polynomials, and periodic systems have also been developed. The inertia of a matrix with respect to a half plane is (denoted by $\text{In}(A)$) is the triplet of the numbers of eigenvalues of $A$ with positive, negative, and zero real parts. The inertia with respect to the unit circle is analogously defined.
In this paper, we briefly survey the stability and inertia theory of matrices, mention a few applications, and describe a nonspectral, implicit matrix equation method for determining the stability and inertia of a matrix. The survey will conclude with a few problems for further research. Earlier surveys on inertia include [17,25]. We emphasize that this survey is about stability and inertia related to matrices only. Inertia results for operators, for general polynomial matrices, and periodic inertia theorems, have not been included. Even for the matrix case, the survey is not claimed to be complete.

For an account of the inertia theory for operators, see [23–25] (and the references therein), and [84,85]. The inertia theory of periodic systems, has been discussed in [17,18,66,106]. For stability and inertia results of matrix polynomials, see [80,81,86–88]. For inertia results related to the algebraic Riccati equations, see [17,29,119].

Since there is a vast literature in this area, omission of some references is inevitable. The author apologizes for such inadvertent omissions.

2. Some facts about Bezoutian, controllability and observability

In this section, we will state some well-known facts about Bezoutian, controllability and observability for convenient use later in the paper.

2.1. The Bezoutian matrix

Let \( f(x) = x^n - a_n x^{n-1} - a_{n-1} x^{n-2} - \cdots - a_2 x - a_1 \) and \( g(x) = b_m x^m - b_{m-1} x^{m-2} - \cdots - b_2 x - b_1 \) be two complex polynomials of degree \( n \) and \( m \), respectively, with \( m \leq n \). Then the complex symmetric matrix \( B = (b_{ij}) \) defined by the bilinear form

\[
B(f, g) = \frac{f(x)g(y) - f(y)g(x)}{x - y} = \sum_{i,j=0}^{n-1} b_{ij} x^i y^j
\]

is called the Bezoutian matrix associated with \( f(x) \) and \( g(x) \). Let

\[
A = \begin{pmatrix}
0 & 1 & 0 & \cdots & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & \cdots & \cdots & a_n
\end{pmatrix}
\]

be the companion matrix of \( f(x) \). Define
Lemma 2.1 [7,41]. The Bezoutian matrix $B$ defined above is such that

$$ B = U g(A). $$

(2.3)

2.2. Bezoutian as a symmetrizer

By direct matrix multiplication, it is easy to see that $U$ is a symmetrizer of $A$, that is,

$$ U A = A^T U. $$

(2.4)

Again, from (2.3) and (2.4), we have

$$ B A = U g(A) A = U A g(A) = A^T U g(A) = A^T B. $$

Lemma 2.2 [42,43]. The Bezoutian matrix $B$ associated with two polynomials $f(x)$ and $g(x)$ is such that

$$ B A = A^T B, $$

where $A$ is the companion matrix of $f(x)$ in the form (2.1).

For an excellent account of Bezoutian and related results, see [65]. For results on generalized Bezoutians for matrix polynomials, see [1,80,81,86,87]. Householder [70] is another illuminating paper on the Bezoutian.

2.3. Controllability and observability

The two basic concepts in control theory are controllability and observability of a control system.

Definition 2.1. The continuous-time linear time-invariant system

$$ \begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t), \\
y(t) &= Cx(t)
\end{align*} $$

(2.5)

is said to be controllable if starting from any initial state $x(0)$, the system can be driven to any final state $x_f$ in some finite time $t_f$, choosing the input vector $u(t)$, $0 \leq t \leq t_f$, appropriately.

Observability is a dual concept of controllability.
**Definition 2.2.** The system (2.5) is said to be observable if there exists \( t_1 > 0 \) such that the initial state \( x(0) \) can be uniquely determined from the knowledge of \( y(t) \) for all \( t, 0 \leq t \leq t_1 \).

Since the matrix \( C \) does not have any role in the definition of controllability, the controllability of (2.5) is often referred to as the controllability of the pair \((A, B)\). Similarly, since \( B \) does not have any role in the definition of observability, the observability of (2.5) is often referred to as the observability of the pair \((A, C)\).

Some well-known criteria of controllability and observability are now stated in the next two theorems. These criteria will be used later in some of our proofs. The proofs of Theorems 2.1 and 2.2 below can be found in [75].

In the following, \( A \) is \( n \times n \), \( B \) is \( n \times m \) \((m \leq n)\), and \( C \) is \( r \times n \) \((r \leq n)\).

**Theorem 2.1** (Criteria for continuous-time controllability). The pair \((A, B)\) is controllable if and only if any of the following equivalent conditions holds:
1. The controllability matrix
   \[
   C_M = (B, AB, A^2B, \ldots, A^{n-1}B)
   \]
   has rank \( n \).
2. Rank \((A - \lambda I, B)\) = \( n \) for every eigenvalue \( \lambda \) of \( A \).
3. Let \((\lambda, x)\) be an eigenpair of \( A^*\), i.e., \( x^*A = \lambda x^*\), then \( x^*B \neq 0 \).

**Remark 2.1.** Criteria 2 and 3 of Theorem 2.1 are commonly known as the Popov–Belevitch–Hautus (PBH) criteria of controllability [75]. See also [64].

**Theorem 2.2** (Criteria for continuous-time observability). The pair \((A, C)\) is observable if and only if any one of the following equivalent conditions holds:
1. The observability matrix
   \[
   O_M = \begin{pmatrix}
   C \\
   CA \\
   CA^2 \\
   \vdots \\
   CA^{n-1}
   \end{pmatrix}
   \]
   has rank \( n \).
2. The matrix \( \lambda I - \begin{pmatrix} A \\ C \end{pmatrix} \) has rank \( n \) for every eigenvalue \( \lambda \) of \( A \).
3. Let \((\lambda, y)\) be an eigenpair of \( A\), i.e., \( Ay = \lambda y\), then \( Cy \neq 0 \).

**3. Matrix stability theory**

In this section, we briefly describe the classical stability results of the systems of the differential and difference equations (1.1), and (1.2).
3.1. Stability of systems of differential equations

A well-known criterion of stability of the system (1.1) is:

**Theorem 3.1.** A necessary and sufficient condition for the system (1.1) to be asymptotically stable is that all the eigenvalues of the matrix $A$ have negative real parts.

**Proof.** It is well known that the general solution of the system (1.1) is given by

$$x(t) = e^{At}x_0.$$  

Now, assume (without any loss of generality) that $A$ is given in Jordan canonical form; that is

$$A = \text{diag}(J_1, \ldots, J_r).$$

Then $e^{At} = \text{diag}(e^{J_1 t}, e^{J_2 t}, \ldots, e^{J_r t})$. Again, $e^{A t} = e^{\lambda_k t} T$, where $T$ is a nonsingular upper triangular matrix and $\lambda_k$ is an eigenvalue of $A$ (see [51]). Let $\lambda_k = \alpha_k + i\beta_k$, then $e^{\lambda_k t} \to 0$ if and only if $\alpha_k < 0$. Thus, $e^{At} \to 0$, as $t \to \infty$, if and only if all the eigenvalues of $A$ have negative real parts. □

**Definition 3.1.** A matrix $A$ is called a stable matrix if all of the eigenvalues of $A$ have negative real parts.

Knowing that the system (1.1) is asymptotically stable if and only if $A$ is a stable matrix, a connection of the stability theory with a matrix equation may be established via the matrix $X$ defined by

$$X = \int_0^\infty e^{A^* t} M e^{A t} \, dt,$$  

(3.1)

where $M$ is an arbitrary positive definite matrix.

Note that when $A$ is stable, $e^{At} \to 0$ at $t \to \infty$; thus $X$ is defined, and it is easy to see (see the proof of Theorem 3.2 below) that $X$ satisfies the equation

$$XA + A^* X = -M.$$  

(3.2)

This brings us to the matrix version of the Lyapunov stability theory. The following theorem (Theorem 3.2) is known as the **Lyapunov stability theorem**. Lyapunov’s original formulation dealt with the stability of a system of nonlinear differential equations. The matrix formulation in the linear case, as stated in Theorem 3.2, possibly first appeared in [57, Vol. II].

The matrix equation (3.2), as well as its dual

$$AX + XA^* = -M$$  

(3.3)

are called **Lyapunov equations**.
3.2. Lyapunov stability

Several proofs of the Lyapunov stability theorem (Theorem 3.2) are available in the literature [e.g., 11, 59, Vol. II; 63]. The proof given here is along the line in [11].

**Theorem 3.2** (The Lyapunov stability theorem). The system (1.1) is asymptotically stable if and only if, for any Hermitian positive definite matrix $M$, there exists a unique Hermitian positive definite matrix $X$ satisfying the Lyapunov equation (3.2).

**Proof.** We first show that the matrix defined by (3.1) is a unique Hermitian positive definite solution of (3.2) if $A$ is stable.

Substituting the expression of $X$ from (3.1) in the Lyapunov equation (3.2), we obtain

$$
XA + A^*X = \int_0^\infty e^{A^t}Me^{At} \, dt + \int_0^\infty A^*e^{A^t}Me^{At} \, dt
= \int_0^\infty \frac{d}{dt}(e^{A^t}Me^{At}) \, dt = e^{A^t}Me^{At} \bigg|_0^\infty .
$$

Since $A$ is stable, $e^{A^t} \to 0$ as $t \to \infty$. Thus $XA + A^*X = -M$, showing that $X$ defined by (3.1) satisfies the Lyapunov equation (3.2).

To prove that $X$ is unique, assume that there are two solutions $X_1$ and $X_2$.

Then

$$
A^*(X_1 - X_2) + (X_1 - X_2)A = 0,
$$

which implies that

$$
e^{A^t}(A^*(X_1 - X_2) + (X_1 - X_2)A)e^{At} = 0,$$

or

$$
\frac{d}{dt} \left[ e^{A^t}(X_1 - X_2)e^{At} \right] = 0.
$$

Hence, $e^{A^t}(X_1 - X_2)e^{At}$ is a constant matrix for all $t$. Evaluating this expression at $t = 0$ and $t = \infty$, we conclude that $X_1 - X_2 = 0$. Furthermore, a unique solution $X$ of (3.2) must be Hermitian.

To show that $X$ is positive definite, we have to show that $u^*Xu > 0$ for any nonzero vector $u$.

We can write

$$
u^*Xu = \int_0^\infty u^*e^{A^t}Me^{At}u \, dt.
$$

Since the exponential matrices $e^{A^t}$ and $e^{At}$ are both nonsingular and $M$ is positive definite, we conclude that $u^*Xu > 0$.

We now prove the converse; that is, we prove that if $X$ is a Hermitian positive definite solution of the Lyapunov equation (3.2), then $A$ is stable.
Let $\lambda, x$ be an eigenpair of $A$. Then premultiplying the Lyapunov equation (3.2) by $x^*$ and postmultiplying it by $x$, we obtain

$$x^* X A x + x^* A^* X x = \lambda x^* X x + \bar{\lambda} x^* X x = (\lambda + \bar{\lambda}) x^* X x = -x^* M x.$$ 

Since $M$ and $X$ are both Hermitian positive definite, we have $\lambda + \bar{\lambda} < 0$ or $\text{Re}(\lambda) < 0$. \hfill \Box

The Lyapunov equations also arise in many other important control theoretic applications. In many of these applications, the right-hand side matrix $M$ is positive semidefinite, rather than positive definite. The typical cases are $M = BB^*$ or $M = C^* C$, where $B$ and $C$ are, respectively, the input and output matrices. Some of the important control theoretic applications that give rise to Lyapunov equations of the above type include robust stabilization, computation of $H_2$ and $H_\infty$ norms, balanced-realization and model reduction [52,59]. We now state two results on the existence of positive definite solutions of the Lyapunov equations with positive semidefinite right-hand side matrices of the above types. See [1,82,83] for these and other related results.

**Theorem 3.3.** Let $A$ be a stable matrix. Then the Lyapunov equation

$$XA + A^* X = -C^* C$$

(3.4)

has a unique Hermitian positive definite solution $X$ if and only if $(A, C)$ is observable.

**Proof.** We first show that the observability of $(A, C)$ and stability of $A$ imply that $X$ is positive definite.

Since $A$ is stable, the unique Hermitian solution $X$ of Eq. (3.4) is given by

$$X = \int_0^\infty e^{A t} C^* C e^{A^* t} \, dt.$$ 

If $X$ is not positive definite, then there exists a nonzero vector $x$ such that $X x = 0$. In that case

$$\int_0^\infty \|Ce^{A t} x\|^2 \, dt = 0.$$ 

Hence $Ce^{A t} x = 0$ for every $t$. Evaluating $Ce^{A t} x = 0$ and its successive derivatives at $t = 0$, we obtain $C A^t x = 0$, $i = 0, 1, \ldots, n - 1$. Since $(A, C)$ is observable, this implies that $x = 0$ (Theorem 2.2). Thus we have a contradiction. So, $X$ is positive definite.

Next, we prove the converse. The proof is again by contradiction.

Since $(A, C)$ is not observable, there is an eigenvector $x$ of $A$ such that $C x = 0$ (Theorem 2.2). Let $\lambda$ be the corresponding eigenvalue. Then from Eq. (3.4), we have

$$(\lambda + \bar{\lambda}) x^* X x = -\|C x\|^2 = 0.$$ 

Since $A$ is stable, $\lambda + \bar{\lambda} < 0$. Thus $x^* X x = 0$. But $X$ is positive definite, so $x$ must be a zero vector, which is a contradiction. \hfill \Box
Since observability is a dual concept of controllability, the following results can be immediately proved by duality of Theorem 3.3.

**Theorem 3.4.** Let $A$ be a stable matrix. Then the Lyapunov equation

$$AX + XA^* = -BB^*$$

(3.5)

has a unique Hermitian positive definite solution $X$ if and only if $(A, B)$ is controllable.

### 3.3. Stability of systems of difference equations

Consider now the system of difference equations (1.2), with an initial value $x(0) = x_0$.

A well-known mathematical criterion for asymptotic stability of the above system is:

**Theorem 3.5.** The system (1.2) is asymptotically stable if and only if all the eigenvalues of $A$ are inside the unit circle.

**Proof.** The proof follows from the fact that the solution of the system (1.2) is given by $x_k = A^k x_0$, and $A^k \to 0$ if and only if all the eigenvalues of $A$ are inside the unit circle.

**Definition 3.2.** A matrix $A$ having all its eigenvalues inside the unit circle is called a discrete-stable matrix.

Each of the theorems in Section 3.2 has a unit circle counterpart. In this case, the Lyapunov equations $AX + A^* X = -M$ and $AX + XA^* = -M$ are, respectively, replaced by their unit circle analogues: $A^* X A - X = -M$ and $A^* X A^* - X = -M$. These equations are called the *Stein equations*, after the name of Stein [108]. The Stein equations are also known as *discrete Lyapunov equations* in the control literature.

In the following, we state and prove a unit circle analogue of Theorem 3.2. The statements and proofs of the unit circle versions of the Theorems 3.3 and 3.4 are analogous. In fact, the Lyapunov and Stein equations are related via the matrix versions of the well-known transformation (known as the Cayley transformation):

$$s = \frac{z - 1}{z + 1}, \quad z = \frac{1 + s}{1 - s}.$$

It can be shown (see, for example, [111]) that if $C$ has all its eigenvalues inside the unit circle, and $A$ and $X$ are such that

$$A = (C + I)^{-1}(C - I),$$
and \( XA + A^*X = -M \), where \( M \) and \( X \) are positive definite, then \( X - C^*XC \) is positive definite, and vice versa.

**Theorem 3.6** (The Stein stability theorem). The system (1.2) is asymptotically stable if and only if, for any Hermitian positive definite matrix \( M \), there exists a unique Hermitian positive definite matrix \( X \) satisfying the Stein equation

\[
A^*XA - X = -M.
\]

**Proof.** The proof follows from Theorem 3.2 using the Cayley transformation above. An independent proof, similar to that of the Lyapunov Stability Theorem, can also be given. This proof follows by noting that the explicit expression of a unique solution \( X \) of (3.6) is given by \( X = \sum_{k=0}^{\infty} (A^*)^k MA^k \).

### 4. The inertia theory of matrices

In this section, we will briefly review inertia theorems with respect to the half planes and the unit circle. The inertia and stability theorems for more general regions will be discussed in the next section. We first formally define the inertia and the unit circle inertia.

**Definition 4.1.** The inertia of a matrix \( A \) of order \( n \), denoted by \( In(A) \), is the triplet \((\pi(A), \nu(A), \delta(A))\), where \( \pi(A), \nu(A), \) and \( \delta(A) \) are, respectively, the number of eigenvalues of \( A \) with positive, negative, and zero real parts, counting multiplicities.

Note that \( \pi(A) + \nu(A) + \delta(A) = n \), and \( A \) is a stable matrix if and only if \( In(A) = (0, n, 0) \). The inertia, as defined above, is the half plane inertia. The inertia with respect to the other regions of the complex plane can similarly be defined. **Unless otherwise stated, by the term “inertia” of a matrix, we will mean the half plane inertia.**

**Remark 4.1.** The term inertia of a nonhermitian matrix \( A \), as defined in Definition 4.1, seems to have been coined by Ostrowski and Schneider [95].

**Definition 4.2.** The unit circle inertia is defined by the triplet \((\pi_0(A), \nu_0(A), \delta_0(A))\), where \( \pi_0(A), \nu_0(A), \) and \( \delta_0(A) \), are, respectively, the number of eigenvalues of \( A \) outside, inside, and on the unit circle. It will be denoted by \( In_0(A) \).

### 4.1. The Sylvester law of inertia

A classical law on the inertia of a Hermitian matrix \( A \) is the Sylvester law of inertia. For a proof, we refer the readers to the book by Horn and Johnson [69] or Cain [25].
Theorem 4.1. Let $A$ be a Hermitian matrix and $P$ be a nonsingular matrix. Then
\[ \text{In}(A) = \text{In}(PAP^*). \quad (4.1) \]

Remark 4.2. Cain [25] has obtained some generalizations of the above Sylvester theorem to a class of normal matrices. He has also recently reported extensions of the Sylvester theorem to some other classes of nonhermitian matrices at the International Linear Algebra Society Meeting, 1998 [26].

The importance of the Sylvester law of inertia. Using the Sylvester law of inertia, the inertia of a given Hermitian matrix $A$ can be computed in terms of the inertia of the diagonal matrix $D$ associated with its triangular factorization. Thus, if $A$ is $n \times n$ and Hermitian and has the triangular factorization $A = LDL^*$, where $L$ is a nonsingular lower triangular matrix, and $D$ is a diagonal matrix with $p$ positive, $q$ negative, and $r$ zero diagonal entries ($p + q + r = n$), then by the Sylvester law of inertia,
\[ \text{In}(A) = (p, q, r). \]

We now turn to the inertia theory for nonhermitian matrices.

4.2. Inertia theory for the Lyapunov and Stein equations

The Sylvester law of inertia and the matrix formulation of the Lyapunov stability theory (Theorem 3.2) seem to have made a significant impact on the development of the nonhermitian inertia theorems. Indeed, Schneider [104] has remarked that “Gantmacher’s reformulation had a deep influence on the inertia theory of matrices as developed in 1960’s and subsequently”.

In this section and elsewhere in the paper, $M > 0$ ($\geq 0$) means that the matrix $M$ is Hermitian positive definite (positive semidefinite).

Theorem 4.2 (The main inertia theorem).

(i) A necessary and sufficient condition that there exists a Hermitian matrix $X$ such that
\[XA + A^*X = M > 0 \quad (4.2)\]
is that $\delta(A) = 0$.

(ii) If $X$ is Hermitian and satisfies (4.2), then $\text{In}(A) = \text{In}(X)$.

Remark 4.3. [Historical development of the main inertia theorem] The main inertia theorem (Theorem 4.2), as it appears above, is due to Ostrowski and Schneider [95]. However, Part (ii) of the Theorem was proved independently by Taussky [110] in the special case when $\Delta(A) = \prod_{i,j=1}^n (\lambda_i + \bar{\lambda}_j) \neq 0$, the $\lambda_i$ being the eigenvalues of $A$; that is, when $X$ is a unique solution of the Lyapunov equation, which is necessarily Hermitian. Krein proved a result in Banach space which in finite dimensions is equivalent to Part (i) (see [40]). Wielandt [114] gave a result which is equivalent to Part (ii).

The following corollary of Theorem 4.2 is immediate.

Corollary 4.1. A is a stable matrix if and only if there exist a negative definite matrix $X$ such that

$$XA + A^*X = M > 0.$$  \hspace{1cm} (4.3)

Remark 4.4. Recovery of the Lyapunov stability theorem and the Sylvester law of inertia from the main inertia theorem] Corollary 4.1 together with the fact that when $A$ is stable, the Lyapunov equation (3.2) has a unique solution [11], yield the Lyapunov stability theorem (Theorem 3.2). A proof of the Sylvester law of inertia using the MIT has been given by Cain [25].

The following unit circle analogue of the Main Inertia Theorem was mentioned by Taussky [111] (Remark 2 in that paper), but not proved. The result is implicit in [67]. A formal proof of the theorem appears in [116]. For an operator version of this theorem, see [24, 25]. See also [115].

Theorem 4.3. There exists a Hermitian $X$ such that

$$A^*XA - X = M > 0$$  \hspace{1cm} (4.4)

if and only if $\delta_0(A) = 0$. In this case $I_{n_0}(A) = I_{n}(X)$.

While the Main Inertia Theorem (Theorem 4.2) and its unit circle counterpart (Theorem 4.3) generalize the Lyapunov and Stein stability theorems (Theorems 3.2 and 3.6) and are important contributions to the literature in their own rights, their practical uses are restricted. As we have seen before and will see again later that many control-theoretic applications give rise to Lyapunov matrix equations with positive semidefinite right-hand sides, and thus Theorems 4.2 and 4.3 are not applicable in these situations.

The next inertia theorem proved by Carlson and Schneider [35] deals with the positive semidefinite case.

Theorem 4.4 [35]. Let $\delta(A) = 0$, and let $X$ be a nonsingular Hermitian matrix such that $XA + A^*X = M \succeq 0$, then $I_{n_0}(A) = I_{n}(X)$.

A unit circle analogue of Theorem 4.4 appears in [46].

Theorem 4.5. Let $\delta_0(A) = 0$, and let $X$ be a nonsingular Hermitian matrix such that $A^*XA - X = M \succeq 0$. Then $I_{n_0}(A) = I_{n}(X)$.

The applicability of Theorem 4.4 requires that $\delta(A) = 0$ and similarily, Theorem 4.5 requires that $\delta_0(A) = 0$. However, we will see in the following theorems that
these conditions can be replaced by some appropriate controllability conditions. Specifically, it will be shown that the controllability of \((A^*, M)\), or equivalently, the observability of \((A, M)\), implies that \(\delta(A) = 0\) [36,118].

The other related inertia results involving controllability and inertia can be found in [28,33,34]. See also [32] for another result on the controllability of \((A^*, M)\) and the semidefiniteness of \(M\).

**Theorem 4.6** [36,117,118]. Let \(X\) be a nonsingular Hermitian matrix such that \(XA + A^*X = M \succeq 0\), and let \((A^*, M)\) be controllable. Then \(\delta(A) = 0\), and \(\text{In}(A) = \text{In}(X)\).

**Proof.** Suppose that \((A^*, M)\) is controllable, but \(\delta(A) \neq 0\). Since \(\delta(A) \neq 0\), there is an eigenvalue \(\lambda\) of \(A\) such that \(\lambda + \lambda = 0\). Let \(x\) be an eigenvector of \(A\) corresponding to \(\lambda\). Then

\[
x^*Mx = x^*(XA + A^*X)x = (\lambda + \lambda)x^*Xx = 0.
\]

This is, according to the eigenvector criterion of controllability (Theorem 2.1), in contradiction to the assumption that \((A^*, M)\) is controllable. Inertia conclusion of Theorem 4.6 now follows from the Carlson–Schneider theorem (Theorem 4.4).

**Remark 4.5.** Though we have stated Theorem 4.6 in a way so that uniformity with the statement of Theorem 4.4 is maintained, the assumption on the nonsingularity of \(X\) in Theorem 4.6 can be dropped; because, it can be shown that the controllability of \((A^*, M)\) also implies that \(\delta(X) = 0\). Indeed, a slightly stronger result can be proved for the quadratic equation

\[
XA + A^*X = X^*B^*B X.
\]

The above equation was studied by Carlson and Datta [29] under the assumption of the controllability of \((A, B^*)\), allowing \(X\) to be nonhermitian.

It was shown in Theorem 4 of that paper that \(X\) is nonsingular if and only if \((A^*, X^*B)\) is controllable.

The unit circle version of Theorem 4.5 was obtained by Wimmer and Ziebur [120].

**Theorem 4.7** [120]. Let \(X\) be a nonsingular Hermitian matrix such that \(A^*XA - X = M \succeq 0\), and let \((A^*, M)\) be controllable. Then \(\delta_0(A) = 0\), and \(\text{In}_0(A) = \text{In}(X)\).

It is natural to ask what happens if the pair \((A^*, M)\) is not controllable. In this case the controllability subspace \(L(A^*, M) = \text{span}(M, A^*M, \ldots, (A^*)^{n-1}M)\) has rank less than \(n\), say \(l\). Snyders and Zakai [109] proved an inertia result in the case \(X > 0\), and Loewy [90] established inertia inequalities for any Hermitian \(X\), assuming that the controllability subspace \(L(A^*, M)\) has rank \(l < n\).
**Theorem 4.8** [109]. Let $X > 0$ be such that $XA + A^*X = M \succeq 0$, then $\nu(A) = 0, \delta(A) = n - l, \pi(A) = l$.

**Remark 4.6.** A proof of Theorem 4.8 using Theorem 4.6 appears in [118].

**Theorem 4.9** [90]. Let $X$ be a Hermitian matrix such that $XA + A^*X = M \succeq 0$. Then

$$\|\pi(A) - \pi(X)\| \leq n - l,$$

$$|\nu(A) - \nu(X)| \leq n - l.$$

### 4.3. Generalized stability theory

Since the Lyapunov and Stein’s stability theorems deal, respectively, with half planes and the unit circle, it is only natural to investigate whether these theorems can be generalized with respect to more general regions of the complex plane. For example, one could ask the following question:

Given an $n \times n$ matrix $A$, a Hermitian positive definite or semidefinite matrix $K$, and the region $R = \{ z \mid r(z, \bar{z}) > 0 \}$, where $r(\lambda, \mu)$ is a polynomial with complex coefficients,

$$r(\lambda, \mu) = \sum_{i,j=1}^{n} d_{ij}\lambda^{i-1}\mu^{j-1}, \quad (4.5)$$

what conditions guarantee the existence of a Hermitian positive definite solution $X$ of the equation

$$\sum_{i,j} d_{ij}A^{i-1}X(A^*)^{j-1} = K.$$ 

More generally, how does the inertia of $X$ provide information on the eigenvalue distribution of $A$ inside and outside the region $R$?

Several attempts have been made over the years to answer the above and related questions. These include the contributions by Schneider [103], Kalman [77], Gutman and Jury [62], Gutman [61], Kharitonov [78], Djaferis and Mitter [53], Mazko [94], Howland [71], Chen [37] and Hill [67]. Some resulted in partial success and others in failure. The first such success was due to Schneider who proved the following theorem which deals with a matrix equation more general than above, but the matrix $D = (d_{ij})$ is very special.

**Theorem 4.10** [103]. Let $C, A_1, A_2, \ldots, A_s$ be the complex matrices of order $n$ which are simultaneously triangulable. Let
Then the following are equivalent:

(i) For any $M > 0$, there exists a unique $X > 0$ such that $T(X) = M$.

(ii) \[ |\alpha_i|^2 - \sum_{k=1}^s |v_i^{(k)}|^2 > 0, \quad i = 1, \ldots, n, \] (4.7)

where $\alpha_i, v_i^{(k)}, i = 1, \ldots, n$ and $k = 1, \ldots, s$, are the eigenvalue of $C$ and $A_k$, under a natural correspondence.

(iii) There exists $X > 0$ such that $T(X) > 0$.

Remark 4.7 (Recovery of Lyapunov’s and Stein’s Stability Theorems from Schneider’s theorem). In Theorem 4.10, if one sets $C = A^* + I$, $s = 2$, $A_1 = A^*$, $A_2 = I$, then one obtains the Lyapunov stability theorem (Theorem 3.2). Similarly, if one sets $C = I$, $s = 1$, $A_1 = A^*$, then one obtains the Stein stability theorem (Theorem 3.6).

The following technical result due to David Carlson has appeared in [67]. It will be used later in discussing relationships between various theorems.

Theorem 4.11 [Carlson (1969)]. Let $A_1, \ldots, A_s$ be $n \times n$ simultaneously triangularisable complex matrices whose eigenvalues $\lambda_k^{(i)}, k = 1, \ldots, n, i = 1, \ldots, s$, are under a natural correspondence. Let $X$ be a Hermitian matrix of order $n$. Let $D = (d_{ij})$ be a Hermitian matrix of order $s$ with eigenvalues $\delta_1, \ldots, \delta_s$, and let $T'(X) = \sum d_{ij} A_i X A_j^*$. Then there exist simultaneously triangularisable matrices $B_1, B_2, \ldots, B_s$ of order $n$ with eigenvalues $\mu_k^{(i)}, k = 1, \ldots, n, i = 1, \ldots, s$, under the same correspondence such that

\[ T'(X) = \sum_{i=1}^s \delta_i B_i X B_i^* , \]

and

\[ \sum_{i,j=1}^s d_{ij} \lambda_k^{(i)} (\bar{\lambda}_k)^j = \sum_{i=1}^s \delta_i |\mu_k^{(i)}|^2, \quad k = 1, \ldots, n. \]

Several researchers (e.g. [53,62,77,78]), being unaware of the Schneider theorem (Theorem 4.10), proved results which are, in some sense, special cases of this theorem (for details, see later in this section). Below we quote Kharitonov’s result.
which includes Kalman’s and others, and then exhibit a relationship of his theorem (Theorem 4.12) with Theorem 4.10.

**Definition 4.3.** The signature of a Hermitian matrix $D$, denoted by signature($D$), is defined to be $\pi(D) - \nu(D)$.

**Theorem 4.12** [78]. Let $D = (d_{ij})$ be an $s \times s$ Hermitian matrix such that rank ($D$) + signature ($D$) = 2. Then the following are equivalent.

(i) For any $M > 0$, there exists a unique $X > 0$ such that

$$T''(X) = \sum_{i,j=1}^{s} d_{ij} A^{j-1} X (A^*)^{i-1} = M.$$

(ii) $\sum_{i,j=1}^{s} d_{ij}(\alpha_k)^{i-1}(\bar{\alpha}_k)^{j-1} > 0$, where $\alpha_k, k = 1, \ldots, n$, are the eigenvalues of $A$.

**Remark 4.8.** Kalman [77] proved Theorem 4.12 under the assumption that rank ($D$) = 2, signature ($D$) = 0. Note that this assumption implies $\pi(D) = \nu(D) = 1$, and conversely.

**Remark 4.9** (Relationship between Schneider’s and Kharitonov’s theorems). Schneider’s result is more general than Kharitonov’s in the sense that the matrices considered there are more general. It deals with arbitrary simultaneously triangulable matrices $C, A_1, \ldots, A_s$, whereas Kharitonov’s result deals with (the obviously simultaneously triangulable) matrices $I, A, \ldots, A'$. However, it is less general than Kharitonov’s result in the sense that the matrix $D = (d_{ij})$ in Schneider’s theorem is a diagonal matrix. For example, in the case $D = \text{diag}(1, -1, \ldots, -1)$, Kharitonov’s theorem can be recovered from Schneider’s theorem by taking

$$C = I, \quad A_k = A^k, \quad k = 1, \ldots, s - 1; \quad A_s = 0.$$

(Note that rank ($D$) + signature ($D$) = 2.)

**Remark 4.10** (Recovery of other stability theorems from Kharitonov’s theorem). Kharitonov’s theorem (and therefore Schneider’s theorem) include not only the Lyapunov and Stein stability theorems (because, for each of these theorems, $\pi(D) = \gamma(D) = 1$), but several other generalized stability theorems. For example, some stability theorems due to Gutman [60], which are more general than the Lyapunov and Stein stability theorems, include half planes, ellipses, the unit circle, parabolas, certain hyperbolas, strips, sectors, and more. These theorems can be recovered from Kharitonov’s theorem, because in each case rank($D$) + signature($D$) = 2.
4.4. A generalized stability and inertia theorem

Schneider’s Kharitonov’s, and others by Gutman, etc. are generalized stability theorems. On the other hand, Hill [67] has obtained both generalized stability and generalized inertia theorems under a stronger hypothesis; namely, the quasi-commutativity of \( \{A_i\} \).

Below we state Hill’s results using his terminology and then show how the MIT (Theorem 4.2), its Stein analogue (Theorem 4.3), and a part of Schneider’s theorem, can be recovered as special cases.

**Definition 4.4.** The complex matrices \( A_1, A_2, \ldots, A_s \) are said to be quasi-commutative if each of \( A_k \) commutes with \( A_i A_j - A_j A_i \) for \( i, j \in \{1, 2, \ldots, s\} \).

**Theorem 4.13** [67]. Let \( A_1, \ldots, A_s \) be quasi-commutative matrices whose eigenvalues \( \lambda_k^{(1)}, \lambda_k^{(2)}, \ldots, \lambda_k^{(r)} (k = 1, \ldots, n) \) are under a natural correspondence and let \( D = (d_{ij}) \) be Hermitian of order \( s \). Define \( T^{(i)}(X) = \sum_{i,j=1}^{s} d_{ij} A_i X A_j^* \).

(a) The following are equivalent:

(i) There exists a matrix \( X > 0 \) such that \( T^{(i)}(X) > 0 \).

(ii) \( \sum_{i,j=1}^{s} d_{ij} \lambda_k^{(i)} \lambda_k^{(j)} > 0 \), \( k = 1, 2, \ldots, n \).

(b) The following are equivalent:

(iii) There exists a matrix \( X \) such that \( T^{(i)}(X) > 0 \).

(iv) \( \sum_{i,j=1}^{s} d_{ij} \lambda_k^{(i)} \lambda_k^{(j)} \neq 0 \), \( k = 1, 2, \ldots, n \).

(c) Furthermore, if \( \pi(D) \leq 1, v(D) \leq 1 \), and \( X \) is a Hermitian matrix such that \( T^{(i)}(X) > 0 \), then \( \ln(X) = (\pi', v', \delta') \), where \( \pi', v'^{\prime}, \) and \( \delta' \) are, respectively, the number of positive, negative, and zero values of \( \sum_{i,j=1}^{s} d_{ij} \lambda_k^{(i)} \lambda_k^{(j)} \).

Notes:

1. The equivalence (i) \( \Leftrightarrow \) (ii) in (a) has the form of (ii) \( \Leftrightarrow \) (iii) of Theorem 4.10, and may be considered to be generalized stability theorem. The rest of Theorem 4.13 is a generalized inertia theorem.

2. In [104] it was noted (Theorem 7) that Theorem 4.13(a) does not require quasi-commutativity, but rather the (weaker) common eigenvector property: for every distinct sequence \( \{\lambda_k^{(1)}, \lambda_k^{(2)}, \ldots, \lambda_k^{(r)}\} \) of corresponding eigenvalues of the matrices \( A_1, A_2, \ldots, A_s \), there is a common eigenvector. This is in fact also true for (b) of Theorem 4.13.

3. Hill [67] has shown by an example that if \( \pi(D) > 1 \), then his conclusions of (a) and (b) do not hold for arbitrary sequences \( A_1, A_2, \ldots, A_s \) of simultaneously triangulable matrices.
Remark 4.11 (Recovery of the main inertia theorem and its Stein’s analogue from Hill’s theorem). Take \( A_2^* = A, A_1 = I \), and

\[
D = \begin{pmatrix}
0 & 1 \\
1 & 0 
\end{pmatrix},
\]

then Hill’s theorem (Theorem 4.13) becomes the Main Inertia Theorem. On the other hand, if

\[
D = \begin{pmatrix}
1 & 0 \\
0 & -1 
\end{pmatrix},
\]

then Hill’s Theorem becomes the Stein analogue of the MIT (Theorem 4.3). Of course, the Lyapunov and the Stein stability theorems also follow from Hill’s Theorem, since they are, respectively, special cases of the MIT and Theorem 4.3.

Remark 4.12 (Relationship between Schneider’s, Hill’s, and Theorem (4.12)). Hill stated that his conclusion (a) holds when \( \pi(D) = 1 \) and \( A_1, \ldots, A_s \) are simultaneously triangulable. Theorem 4.10, together with Theorem 4.11, allows an easy proof of this. On the other hand, Theorem 4.13 can be used with Theorem 4.11 to prove (ii)\( \Leftrightarrow \) (iii) of Theorem 4.10 whenever \( A_1, \ldots, A_s \) satisfy the common eigenvector property. In [104], Theorem 4.10 is used with Theorem 4.11 to prove Kharitonov’s Theorem. Hill also stated that his conclusions (b) and (c) hold whenever \( \pi(D) \leq 1, \nu(D) \leq 1, \) and \( A_1, \ldots, A_s \) are simultaneously triangulable. The MIT and Sylvester’s Law of Inertia can be used with Theorem 4.11 to provide a proof of this claim.

Note that both Schneider’s theorem and Kharitonov’s theorem deal with matrix equations with arbitrary positive definite right-hand side matrices, but Hill’s theorem does not.

An open question. Since the matrix \( D \) in each of Kharitonov’s theorem, Schneider’s theorem (implicitly) and Hill’s theorem is such that \( \text{rank}(D) + \text{signature}(D) = 2 \), an obvious question is: What is the largest region satisfying the condition \( \text{rank}(D) + \text{signature}(D) = 2 \)?

Chojnowski and Gutman [38] have proved that the largest family of regions in the linear matrix equation is the family of \( M \)-transformable regions (for the definition of this region, see [38,94]). This includes the case that \( \text{rank}(D) + \text{signature}(D) = 2 \). On the other hand, Kharitonov [78] has given the following example to show that if the above condition does not hold, then Theorem 4.12 may be invalid.

Example 4.1. Let

\[
A = \begin{pmatrix}
1 & 0 \\
0 & 2 
\end{pmatrix}.
\]

Then the matrix equation

\[
9X - 4AX - 4XA^* + A^2X(A^*)^2 = M,
\]

is valid.
where
\[ M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \]
admits a unique Hermitian solution
\[ X = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{2} \end{pmatrix}, \]
which is not positive definite. Note that in this case
\[ D = \begin{pmatrix} 9 & -4 & 0 \\ -4 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \]
which has rank 3 and signature 1. However, if
\[ M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \]
then the solution
\[ X = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \]
is positive definite.

Similar examples have been given earlier by Barnett [8] with respect to an ellipse. Hill [68] has obtained inertia theorems with respect to an arbitrary circle, and certain other curves in the complex plane such as a line, parabola and hyperbola. Djaferis and Mitter’s [53] result treats as special cases several important regions such as the half planes and the shifted half planes. However, it does not cover the important regions ellipse and the unit circle. The attempts by Howland [71] and Chen [37] resulted in failure (see the counter example by Carlson and Hill [33]), probably, because they tried to prove their theorems for every positive definite matrix \( K \) without investigating appropriate conditions on the rank and signature of the matrix \( D \).

We should remark at this point that from the point of view of applications it really does not matter if the above inertia theorems do not hold for every symmetric positive definite \( M \); it is more important to know for which positive definite matrices \( M \) the theorems hold. Identifying just one such \( M \) will help.

We conclude by noting that the generalized inertia problem raised in the beginning of this section has not been satisfactorily settled yet. This problem still remains a topic for further research.

5. Applications of inertia and stability theorems

There are some nice applications of the stability and inertia theorems. A few of them are:

1. Elementary and unified matrix-theoretic proofs of several classical results on root separation of polynomials and matrices [42–46,56,77,96,97, and others].
2. Characterization of \(D\)-stable matrices [6,12–14,31,44].
3. Determination of the stability of second-order differential equations arising in vibration and structural analysis, and the inertia of the associated quadratic pencil [118].
4. Elementary derivation of Wall’s criterion on continued fractions and root location of polynomials [117].
5. Nonspectral approach for computing the inertia of a nonhermitian matrix with respect to several regions of the complex plane [30,47,48].

We will discuss briefly applications 1–4 here. Application 5 will be discussed in the next section.

5.1. New and unified proofs of some classical root separation methods

The problem of counting the numbers of zeros of a polynomial in specified regions of the complex plane is known as the root separation problem. The root separation problem that deals with the half planes is called the Routh–Hurwitz problem, and the one that concerns the unit circle is known as the Schur–Cohn problem. There are many methods for solving the Routh–Hurwitz and Schur–Cohn problems. For an account of these methods, see [57, Vol. II; 79,93].

An outstanding classical algebraic method for these two problems is due to Fujiwara [55]. Indeed, many of the methods developed later can be considered as variants of the Fujiwara method. Fujiwara gave an unified treatment for both problems using Bezoutian and the associated quadratic forms. The theorems containing Fujiwara’s solutions are known as the Routh–Hurwitz–Fujiwara and the Schur–Cohn–Fujiwara Theorems [42]. In [45], it was also shown how the well-known Hankel matrices of Markov parameters may be employed to solve the above problems. Theorems 5.1 and 5.2 of the above paper are referred to as the Routh–Hurwitz–Markov and the Schur–Cohn–Markov theorems.

In this section we show how proofs of the Fujiwara methods can be simplified and unified using the inertia theorems of Section 4, and remark how the other root separation methods (such as those contained in the Routh–Hurwitz–Markov and the Schur–Cohn–Markov theorems) can be treated in a similar way. The original proofs of these and other theorems involve quadratic forms and concepts from functional theory.

**Theorem 5.1** (The Routh–Hurwitz–Fujiwara theorem). Let \(f(x)\) be a given complex polynomial of degree \(n\), and let \(B = (b_{ij})\), be the Bezoutian matrix of \(f(x)\) and \(\bar{f}(-x)\), and let the Hermitian matrix \(F = (f_{ij})\) be defined by

\[
f_{ij} = (-1)^i b_{ij}, \quad i, \ j = 0, 1, 2, \ldots, n - 1. \tag{5.1}
\]

Then, whenever \(F\) is nonsingular, the numbers of zeros of \(f(x)\) in right and left half planes are, respectively, equal to the numbers of negative and positive eigenval-
ues of $F$; in particular, $F$ is positive definite if and only if all the zeros of $f(x)$ are in the left half plane.

**Proof.** From (5.1) we have

$$ F = DB = D^* B, \tag{5.2} $$

where $D = \text{diag}(1, -1, 1, -1, \ldots, (-1)^{n-1})$. If $A$ is the companion matrix of $f(x)$, given by (2.1), then

$$ FA + A^* F = D^* B A + A^* D^* B = D^* A^T B + A^* D^* B \quad \text{(since $BA = A^T B$)}$$

$$ = (D^* A^T + A^* D^*) B \tag{5.3}$$

It is trivial to see, by direct computation, that $DA + \tilde{A} D$ is a matrix whose first $(n - 1)$ rows are zero and the last row is $-b_n = -e_n B$. So, from (5.3), we get

$$ FA + A^* F = -b_n^* e_n B = -B^* e_n^* e_n B = -\tilde{B} e_n^T e_n B \leq 0. \tag{5.4} $$

Theorem 5.1 is now proved by applying to (5.4) either Theorem 4.4 or Theorem 4.6. This is seen as follows:

(i) Since $(A, e_n^T)$ is controllable, $BA = A^T B$, and $B$ is nonsingular (the nonsingularity of $F = DB$ implies the nonsingularity of $B$), it then follows that $(A^*, \tilde{B} e_n^T)$ is controllable. Theorem 4.6 now can be applied to (5.4) to conclude the assertions of Theorem 5.1.

(ii) Since $B$ is nonsingular it then follows that $\tilde{f}(-A)$ is nonsingular (by Lemma 2.1). This means that $\delta(A) = 0$. Theorem 4.4 now can be applied to (5.4) to conclude the assertions of Theorem 5.1. \qed

**Remark 5.1.** A Theorem of Carlson and Datta [29] (Theorem 3 in that paper), who studied the homogeneous algebraic Riccati equation of the form (5.4) for an arbitrary matrix $A$, can also be applied to (5.4) to obtain Theorem 5.1.

**Theorem 5.2** (The Schur–Cohn–Fujiwara theorem). Let $f(x)$ be a complex polynomial of degree $n$, let $B = (b_{ij})$, be the Bezoutian matrix of $f(x)$ and $g(x) = x^n \tilde{f}(1/x)$, and let the Hermitian matrix $F = (f_{ij})$ be defined by

$$ f_{ij} = \tilde{b}_{i,n-1-j}. \tag{5.5} $$

Then, whenever $F$ is nonsingular, the numbers of zeros of $f(x)$ inside and outside the unit circle are, respectively, equal to the numbers of positive and negative eigenvalues of $F$; in particular $F$ is positive definite if and only if all the zeros of $f(x)$ are inside the unit circle.

**Proof.** From (5.5), we have

$$ F = \tilde{B} P, $$
where $P = (p_{ij})$ is a permutation matrix such that $p_{i,n-i+1} = 1$, $i = 1, 2, \ldots, n$. Let $A$ be the companion matrix (2.1). Then

$$A^*FA - F = A^*BPA - BP$$

$$= \tilde{B} \tilde{AP}A - \tilde{B}P \quad \text{(since } BA = A^T B, \text{ and } B = B^T). \quad (5.6)$$

By direct matrix multiplications, one can very easily verify that $\tilde{A}PA - P$ is a matrix whose first $(n-1)$ rows are zero and, the last row is $(a\bar{\alpha}_1 - 1, \bar{\alpha}_n + a_1\bar{\alpha}_2, \bar{\alpha}_{n-1} + a_1\bar{\alpha}_3, \ldots, \bar{\alpha}_2 + a_1\bar{\alpha}_n) = -b_n$, where $b_n$ is the last row of the associated Bezoutian matrix $B$. So, from (5.6), we have

$$A^*FA - F = -b_n^*b_n = -B^*e_n^*e_nB = -\tilde{B}e_n^Te_nB \leq 0. \quad (5.7)$$

Theorem 5.2 can now be obtained from (5.7) by applying either Theorem 4.5 or Theorem 4.7. This is seen as follows:

The nonsingularity of $B$ implies that $g(A)$ is nonsingular (Lemma 2.1). Since the eigenvalues of $g(A)$ are $\Pi_{i,j} (\lambda_i, \lambda_j, i - 1)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$, it then follows that $A$ does not have an eigenvalue of modulus one. Theorem 4.5 now can be applied.

Alternatively, since $(A^*, \tilde{B}e_n^*)$ is controllable, one can apply Theorem 4.7 to (5.7) to obtain Theorem 5.2. □

Notes:

1. Parks [96,97] first noted the relationship between the Routh–Hurwitz and the Schur–Cohn problems with the respective matrix equations. His proofs, however, were restricted to the stability criteria only. The Carlson–Schneider semidefinite inertia theorem (Theorem 4.4) appeared only a year after the first paper of Parks was published.

2. An elementary proof of the classical Liénard–Chipart [89] criterion of stability has been given by Datta [43], and subsequently by Fuhrman and Datta [56]. Inertia theorems again have played an important role in the proofs of this theorem.


4. The polynomial matrix theorems of Barnett [8] and Datta [41] for root separation problems can be easily derived from Theorems 5.1 and 5.2, taking into account the relationship between Bezoutian matrix and associated polynomial matrix (Lemma 2.1).

5. The proofs of many more existing root separation methods (e.g., the Schwarz method [105], the Anderson–Jury–Mansur method [3], the general eigenvalue
location method of Datta and Datta [47]) can be derived in the spirit of the proofs of Theorems 5.1 and 5.2. It is believed that new methods for the root separation problem, especially in regions other than the half-planes and the unit circle, can still be developed using inertia theory.

6. Several inertia and stability theorems for matrix polynomials have been obtained by Lancaster and Tismenetsky [80,81] Lerer and Tismenetsky [86,87], and others, by generalizing the concept of Bezoutian for matrix polynomials, and then applying the techniques of Theorems 5.1 and 5.2.

5.2. Applications to D-stability

A stable matrix $A$ is called D-stable if $DA$ is stable for all positive diagonal matrices $D$. The concept of D-stability arises in the stability analysis of general equilibrium economic systems (see [4]). Since the problem was first formulated, several papers on the characterization of D-stable matrices have appeared in linear algebra literature. However, a computationally verifiable characterization of D-stability still does not exist. It is well-known [72] that a necessary condition for D-stability is that all principal minors be nonnegative and at least one of each size by positive. We will denote this class of matrices by $P_0^+$.

This condition is, in general, not sufficient. A sufficient condition for D-stability is that there exists a positive diagonal solution matrix to the Lyapunov matrix equation for some positive definite matrix $K$ on the right-hand side [6]. See also [12,13].

Datta [44] identified two classes of D-stable tridiagonal matrices (namely, the Schwarz and the Routh matrices) by constructing in each case a positive diagonal solution matrix $X$ to a Lyapunov equation with a positive semidefinite right-hand side. This result was later generalized by Carlson et al. [31] who gave a complete characterization of D-stability in the case when $A$ is a tridiagonal matrix. This was generalized to arbitrary acyclic matrices by Berman and Hershkowitz [13]. We will present the main results of this paper without the details of the proof. The result rely on the following lemma obtained by these authors.

Lemma 5.1 [31]. Suppose that $A$ is tridiagonal and that the Lyapunov matrix equation

$$XA + A^*X = 2K$$  \hspace{1cm} (5.8)

admits a positive definite Hermitian solution $X$ for some positive semidefinite Hermitian matrix $K$. Then $A$ is stable if and only if

(a) No eigenvector of $X^{-1}S$, where $2S = XA - A^*X$, lies in the null space of $K$.

Furthermore, if $A$ is nonsingular, then the condition (a) can be replaced with:

(b) No eigenvector of $X^{-1}S$, corresponding to a nonzero eigenvalue, lies in the null space of $K$. 

Remark 5.2. Note that if $K$ is actually positive definite, then the null space of $K$ consists only of the zero vector. Thus one direction of the first part of the Lyapunov stability theorem follows immediately from the above lemma.

We need the following definitions to state our result on $D$-stability. Let $A$ be an irreducible tridiagonal matrix. Write $A = H + S$, where $H$ and $S$ are also tridiagonal. Let $S = \text{diag}(S_1, \ldots, S_k)$, where each $S_i$ is either irreducible or zero. Let $A = \text{diag}(A_1, \ldots, A_k)$ be partitioned conformably with $S$.

If $S_j \neq 0$ and $S_{j+1} \neq 0$, we shall call the last diagonal entry of $A_j$ and the first diagonal entry of $A_{j+1}$ transition entries; the diagonal entries which are not transition entries will be called interior entries. We shall call $a_{i,i}a_{i+1,i+1} - a_{i,i+1}a_{i+1,i}$ a transition minor if $a_{i,i}$ and $a_{i+1,i+1}$ are transition entries (and $a_{i,i+1}a_{i+1,i} > 0$).

For a matrix $A$, let $\phi(A) = (i_1, \ldots, i_p)$ be the sequence of indices of diagonal entries which are not zero. Then:

Theorem 5.3 [31]. Let $A \in P^{+}_0$ be irreducible and tridiagonal. Then $A$ is $D$-stable if and only if one of the following holds:

(a) $\phi(A)$ satisfies $i_1 < 3$ or $i_{h+1} - i_h < 3$ for some $h = 1, 2, \ldots, p - 1$ or $i_p > n - 2$, for interior entries.

(b) At least one transition minor is nonzero.

(c) $S_1 = 0$, or $S_k = 0$, or at least two successive $S_j$ vanish.

Proof. We just state here the key idea of the proof of the ‘if’ part: If any of conditions (a), (b) or (c) is satisfied, then for each positive diagonal matrix $D$, there exists a positive diagonal matrix $F$ such that $F(DA)F^{-1} = H + S$, where $H$ and $S$ are also tridiagonal, and satisfies condition (b) of Lemma 5.1. $A$ is thus $D$-stable.

5.3. Applications to stability and inertia of the quadratic matrix pencil

The matrix second-order system (1.3) with real coefficient matrices $M$, $K$, and $D$, arises in a wide variety of practical applications such as in the mechanical vibrations, and structural design analysis. In control theory, it is the fundamental governing equation in the design of large space structures (LSS) (see [5,15]).

The system (1.3) is asymptotically stable if $\|x(t)\| \to 0$ as $t \to \infty$. In terms of the eigenvalues, it then means that the system (1.3) is asymptotically stable if and only if all the eigenvalues of the quadratic pencil (1.4) have negative real parts.

Similarly, by the inertia of the quadratic pencil (1.4) is defined to be the triplet of the numbers of eigenvalues of $P(\lambda)$ with positive, negative, and zero real parts.

The effective numerical methods for the quadratic eigenvalue problem are still not well developed, especially for large and sparse problems that arise in practical
applications. On the other hand, the coefficient matrices $M$, $K$, and $D$ of the pencil $P(\lambda)$ are all symmetric and there are now numerically viable algorithms for large and sparse symmetric eigenvalue problems (see [58]). Furthermore, to compute the inertia of a symmetric matrix, one really does not need to compute the eigenvalues of the matrix; the inertia can be computed in a much more cheaper way by finding its $LDL^T$ decomposition and then applying the Sylvester law of inertia (see Section 6.1). It is thus, natural, to ask the following question:

Can the stability of the system of second-order differential equations (1.3) and the inertia of the associated quadratic pencil (1.4) be determined in terms of the inertia and stability of the symmetric coefficient matrices $M$, $D$, and $K$ of (1.4)?

A classical result on the above problem is the historical Rayleigh criterion of stability. We state the result below and give an elementary proof.

**Theorem 5.4 (Rayleigh).**

(a) Let $M$, $K$, and $D$ be all symmetric and positive definite, then the system (1.3) is asymptotically stable; that is, all $2n$ eigenvalues of the pencil $P(\lambda)$ have negative real parts.

(b) If $M$ and $K$ are symmetric positive definite and $D = 0$, then all the eigenvalues of $P(\lambda)$ are purely imaginary.

**Proof.** Let $\lambda = \alpha + i\beta$ be an eigenvalue of $P(\lambda) = \lambda^2 M + \lambda D + K$, and $x$ be the corresponding eigenvector. Then it was shown in Datta and Rincon (1993) that

$$\alpha = \frac{-|\lambda|^2 D_x}{|\lambda|^2 M_x + K_x},$$

where $L_x$ denotes $x^TLx$.

**Proof of (a):** Since $M$, $K$, and $D$ are symmetric positive definite, then $M_x$, $D_x$, and $K_x$ are all positive; therefore, $\alpha < 0$.

**Proof of (b):** Since $D = 0$, and $M_x$ and $K_x$ are nonzero, then $\alpha = 0$. \qed

The part (b) of the Rayleigh Theorem (Theorem 5.4) says that if $D = 0$, the eigenvalues have all zero real parts. The question naturally arises as to what happens if $D \geq 0$.

The following result, in case of a positive semidefinite $D$, is due to Walker and Schmitendorf [113].

**Theorem 5.5 [113].** Let $M$ and $K$ be symmetric positive definite and $D$ is symmetric positive semidefinite, then the system (1.3) is asymptotically stable if and only if the pair $(\tilde{K}, \tilde{D})$, where $\tilde{K} = M^{-1/2}K M^{-1/2}$, and $\tilde{D} = M^{-1/2}DM^{-1/2}$, is observable.

The following result due to Wimmer [118] is more general and gives information on the inertia of the pencil $P(\lambda)$ as well, under the assumption of positive semidefinite damping.
Define

\[ A = \begin{pmatrix} O & I \\ -M^{-1}K & -M^{-1}D \end{pmatrix}, \quad W = \frac{1}{2} \begin{pmatrix} K & 0 \\ 0 & M \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \]

Then:

**Theorem 5.6** [118]. Let \( M \) and \( K \) be symmetric and nonsingular, let \( D \) be symmetric positive semidefinite and let \( A \) be defined above. If the pair \((A^T, R)\) is controllable, then

(i) The pencil \( P(\lambda) \) has no purely imaginary eigenvalues.

(ii) \( \nu(P(\lambda)) = \pi(M) + \pi(K) \), and \( \pi(P(\lambda)) = \nu(M) + \nu(K) \).

**Proof.** It is an easy computation to see that \( WA + A^TW = -R \leq 0 \). Thus, the proof follows immediately from the Chen–Wimmer inertia Theorem (Theorem 4.6).

**Remark 5.3.** Suppose that the matrices \( M, K, \) and \( D \) are complex. Let \( M \), and \( K \) be Hermitian; and \( M \) be nonsingular. Then, Lancaster and Tismenetsky [81] gave similar inertia result on \( P(\lambda) \), assuming that \( Re(D) = (D + D^*)/2 \) is positive definite. Recently, Bilir and Chicone [16] have given a new proof of their result, and obtained a new result for the case when \( Re(D) \) is positive semidefinite. For other results on the stability and inertia of the pencil (1.4), see [52].

### 5.4. Applications to continued fractions

We present here an elementary matrix theoretic proof of the well-known criterion of Wall, relating the root location of polynomials to continued fractions. The proof uses the Chen–Wimmer inertia theorem and is due to Wimmer [117].

Let \( p(z) = \sum_{k=0}^{n} c_{n-k} z^k, c_0 = 1, n \geq 1 \) be a polynomial with complex coefficients. Define

\[ q(z) = \frac{1}{2} [p(z) - (-1)^n \bar{p}(z)], \]

where \( \bar{p}(z) \) is the polynomial obtained from \( p(z) \) by replacing the coefficient of \( p(z) \) with their complex conjugates.

**Theorem 5.7** (Wall’s criterion). Consider the continued fraction of the form

\[ \frac{q(z)}{p(z)} = \frac{a_1}{|a_1 + ib_1 - z|} + \frac{a_2}{|ib_2 - z|} + \cdots + \frac{a_n}{|ib_n - z|}, \]

where \( a_i, b_i \) are real and \( a_i \neq 0, i = 1, 2, \ldots, n. \) Then the number of roots of \( p(z) \) in the right (left) half plane is equal to the number of positive (negative) elements in the sequence: \( a_1, a_1a_2, \ldots, a_1a_2 \cdots a_n. \)
Proof. The Jacobi matrix $A$ associated with the continued fraction expansion of $q(z)/p(z)$ is

$$A = \begin{pmatrix}
  a_1 + ib_1 & a_2 & 0 & \cdots & 0 \\
-1 & ib_2 & a_3 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & -1 & ib_n
\end{pmatrix},$$

and $p(z)$ is the characteristic polynomial of $A$. It is easy to see that the diagonal matrix

$$D = (a_1, a_1a_2, \ldots, a_1a_2\ldots a_n)$$

satisfies the Lyapunov equation

$$DA + A^*D = 2R,$$

where $R$ is a positive semidefinite matrix. A short computation also shows that the pair $(A^*, R)$ is controllable. Wall’s result now follows from Theorem 4.6.

6. Computational methods for inertia and stability

In this section, we discuss computational approaches for determining the inertia and stability of a matrix $A$. For the sake of computational simplicity (to avoid the use of complex arithmetic), we assume in this section that the matrix $A$ is real.

Definition 6.1. A square matrix $A = (a_{ij})$ is upper Hessenberg if $a_{ij} = 0$ for $i > j + 1$. $A = (a_{ij})$ is lower Hessenberg if $a_{ij} = 0$ for $j > i + 1$. An upper Hessenberg $A$ matrix is unreduced if $a_{i,i-1} \neq 0$ for $i = 2, 3, \ldots, n$. A lower Hessenberg matrix $A$ is unreduced if $a_{i,i+1} \neq 0$ for $i = 1, 2, \ldots, n - 1$.

6.1. Computing the inertia of a symmetric matrix (a nonspectral approach)

If $A$ is symmetric, then the Sylvester law of inertia provides us with an inexpensive and numerically effective method for computing the inertia. A symmetric matrix $A$ admits a triangular factorization

$$A = UDU^T,$$

where $U$ is a product of elementary upper triangular with unit diagonal and permutation matrices, and $D$ is symmetric block diagonal with blocks of order 1 or 2. This is known as diagonal pivoting factorization (see [20–22]). Thus by the Sylvester law of inertia, $In(A) = In(D)$. 

Let $D$ have $p$ blocks of order 1 and $q$ blocks of order 2. Assume that none of the $2 \times 2$ blocks of $D$ is singular. Also, out of the $p$ blocks of order 1, let $p'$ be positive, $p''$ be negative, and $p'''$ be zero (that is, $p' + p'' + p''' = p$). Then,
\[
\pi(A) = p' + q, \\
v(A) = p'' + q, \\
\delta(A) = p'''.
\]
The diagonal pivoting factorization requires only $n^3/6$ flops. It is, thus, twice as efficient as Gaussian elimination process for a nonsymmetric matrix $A$. Furthermore, the process is numerically stable.

6.2. Computing the inertia and testing the stability of a nonsymmetric matrix

The following are the usual computational approaches for determining the stability and inertia of a nonsymmetric matrix $A$:
1. Compute the eigenvalues of $A$ explicitly.
2. Compute the characteristic polynomial of $A$ and then apply the well-known Routh–Hurwitz criterion [93].
3. Solve the Lyapunov equation

\[ XA + A^TX = -C, \]

choosing $C$ conveniently as a positive definite matrix, and then checking if $X$ is positive definite.

The second approach is usually discarded as a numerical approach. This is because, computing the characteristic polynomial of a matrix may be a numerically unstable process [51] or [58]. The process of computing the characteristic polynomial of a matrix $A$ comes in two stages: $A$ is first transformed to an upper Hessenberg matrix $H$ by orthogonal similarity, and then, assuming that $H$, is unreduced, it is further transformed to a companion matrix by nonorthogonal similarity, from where the coefficients of the characteristic polynomial are easily read. The stage 1, that is, the transformation to Hessenberg form, can be accomplished in a numerically stable way using Householder’s or Givens’ method [51] or [58], but stage 2 is, in general, a numerically unstable process. If the transformed Hessenberg matrix has one or more small subdiagonal entries, the corresponding transforming matrix will then be ill-conditioned.

The last approach (the Lyapunov equation approach) is counterproductive. The only numerically effective method for solving the Lyapunov equation is the Schur method of Bartels and Stewart [10]. The method requires transformation of $A$ to real schur form (RSF) which contains the eigenvalues of $A$.

Thus, the only viable way, from a numerical viewpoint, of determining the stability and inertia of a matrix, is to explicitly compute its eigenvalues.
Having said this, let us point out that there exists a computational method due to Carlson and Datta [30] for determining the inertia of a nonsymmetric matrix. The method is direct in the sense that it does not require eigenvalue computations nor it requires solution of a matrix equation. The method is based on the implicit solution of a special Lyapunov equation. Starting from a nonsymmetric matrix $A$, the method constructs a symmetric matrix $X$ such that if $X$ is nonsingular, then $\text{In}(A) = \text{In}(X)$. The matrix $X$ turns out to be a solution of a rank-one positive semidefinite Lyapunov equation which is not explicitly solved. Of course, once the symmetric matrix $X$ is constructed, its inertia can be computed using the Sylvester law of inertia.

We will describe the method below, and then show how some of the semidefinite inertia theorems of Section 4 can be used in the proof of this method. The proof given here is much simpler than the original proof.

An inertia method (Carlson and Datta [29,30])

Step 1. Transform $A$ to a lower Hessenberg matrix $H$ using an orthogonal similarity. Assume that $H$ is unreduced.

Step 2. Construct a nonsingular lower triangular matrix $L$ such that

$$LH + HL = R = \begin{pmatrix} 0 \\ r \end{pmatrix}$$

is a matrix whose first $(n-1)$ rows are zero, starting with the first row $l_1$ of $L$ as $l_1 = (1, 0, \ldots, 0)$.

Step 3. Having constructed $L$, compute the last row $r$ of $R$.

Step 4. Construct now a matrix $S$ such that

$$SH = H^T S,$$

with the last row $s_n$ of $S$ as the last row $r$ of $R$.

Step 5. Compute $F = L^T S$.

Theorem 6.1.
(i) If $F$ is nonsingular, then it is symmetric and $\text{In}(A) = \text{In}(F)$.
(ii) $A$ is stable if and only if $F$ is negative definite.

Proof.

$$FH + H^T F = L^T SH + H^T L^T S$$
$$= L^T H^T S + H^T L^T S$$
$$= (L^T H^T + H^T L^T)S$$
$$= R^T S = r^T r \geq 0. \qed$$

The nonsingularity of $F$ implies the nonsingularity of $S$, and it can be shown (see [48]) that $S$ is nonsingular if and only if $H$ and $-H$ do not have a common eigenvalue. Thus, $F$ is a unique solution of the matrix equation.
and is, therefore, necessarily symmetric. Furthermore, since $H$ and $-H$ do not have a common eigenvalue, then $\delta(H) = 0$. Theorem 4.4 now can be applied to the above matrix equation to obtain Theorem 6.1.

**Remark 6.1 (Computation of $L$).** Once the first row of $L = (l_{ij})$ in step 2 is prescribed, the diagonal entries of $L$ are immediately known. These are: $1, -1, 1, \ldots, (-1)^{n-1}$. Having known these diagonal entries, the $(n(n-1)/2)$ off-diagonal entries $l_{ij}(i > j)$ of $L$ lying below the main diagonal can now be uniquely determined by solving a lower triangular system if these entries are computed in the following order: $l_{21}, l_{31}, l_{32}, \ldots, l_{n1}, l_{n2}, \ldots, l_{n,n-1}$.

**Remark 6.2 (Computation of $S$).** Similar remarks hold for computing $S$ in step 4. Knowing the last row of the matrix $S$, the rows $s_{n-1}$ through $s_1$ of $S$ can be computed directly from the relation $SH = H^TS$.

The algorithm requires only $n^3/2$ operations once the matrix $A$ has been reduced to the Hessenberg matrix $H$. It requires about $5/3n^3$ operations to compute the Hessenberg matrix $H$ from the matrix $A$. Thus a total of about $2n^3$ operations needed to compute the inertia of $A$, compared with about $6n^3$ operations usually needed to compute the eigenvalues of $A$ (see [51, p. 450]). Thus this method is about three times as fast as the implicit double-shift QR method for eigenvalue computations, which is a standard way to compute the eigenvalues of a matrix (see [51] or [58]).

**Notes:**
1. The above algorithm has been modified and made more efficient by Datta and Datta [48]. The modified algorithm uses the matrix-adaptation of the well-known Hyman method for computing the characteristic polynomial of a Hessenberg matrix (see [121]), which is numerically effective with proper scaling.
2. The algorithm has been extended by Datta and Datta [47] to obtain information on the number of eigenvalues of a matrix in several other regions of the complex plane including strips, ellipses, and parabolas.
3. A method of this type for finding distribution of eigenvalues of a matrix with respect to the unit circle has been reported by Lu [91] (an unpublished manuscript (1987).

The semidefinite inertia theorems of Section 4 play an important role in deriving these methods.
6.3. Sensitivity of the inertia problem

In the investigation of backward stability of an inertia algorithm, such as the one just stated, it is important to know the condition number of the inertia problem. Since the inertia problem is basically an eigenvalue problem, it is commonly believed that the sensitivity of these two problems is the same. However, empirical results suggest that this may not be true.

For example, take the well-known 20 \times 20 Wilkinson bidiagonal matrix [121]. A small perturbation of order \(10^{-10}\) in the (20,1)th element of this matrix changes some of the eigenvalues drastically, its inertia, however, remains unchanged.

No formal sensitivity analysis of the inertia problem has yet been done.

7. Some open problems

It is evident from our discussions in this paper that much work has been done on stability and inertia theory; however, there are still areas of inertia that need attention of linear algebraists, numerical linear algebraists, and engineers. We cite a few of them below and discuss very briefly each of them from the perspective of what still remains to be done from our viewpoints.

7.1. Generalized inertia

As stated in Section 4 that the question raised in the beginning of that section still remains unsettled.

Theorems 4.10 and 4.12 and the special cases by Kalman [77] and others are generalized stability theorems. We believe that characterization of generalized inertia via the Hermitian solution of a matrix equation with an arbitrary matrix on the right-hand side is a hard problem. On the other hand, a more tractable and more useful problem, from practical application point of view, is the following:

**Problem 1.** Given an \(n \times n\) matrix \(A\), an \(n \times n\) Hermitian matrix \(D = (d_{ij})\), and the region \(R\) defined by the polynomial (4.5), find a necessary and sufficient condition for the existence of a Hermitian matrix \(X\) such that \(\sum_{i,j=1} d_{ij} A^{i-1} X (A^*)^{j-1}\) is positive definite or semidefinite, and that \(In(X)\) will provide an information on the number of eigenvalues outside, inside, and on the boundary of \(R\).

**Remark 7.1.** Hill’s Theorem (Theorem 4.13) is the most general result proved so far in this direction.
7.2. Singular inertia

The inertia theorems stated in Section 4 are such that they require that the solution matrix $X$ in each case be nonsingular for the result on inertia to hold. This has restricted the applicability of inertia theory to a certain extent. For example, we have stated the Routh–Hurwitz–Fujiwara and Schur–Cohn–Fujiwara Theorems (Theorems 5.1 and 5.2) under the assumption that the Fujiwara matrix in each case is nonsingular and then gave elementary matrix-theoretic proofs of these theorems using the relevant inertia theorems from Section 4. Fujiwara [55], however, discussed the singular cases as well. But Fujiwara’s results in the singular cases could not be proved using inertia theorems; because, the relevant inertia theorems do not exist. Pták and Young [100] have discussed the singular cases of the classical Schur–Cohn criterion, but without any reference to or the use of inertia theory.

The question, therefore, naturally arises:

**Problem 2.** How can the inertia theorems in Section 4 (especially, Theorems 4.4 and 4.6 and their Stein-analogues, Theorems 4.5 and 4.7) be extended to the singular cases?

7.3. Inertia for large, sparse, and structured matrices

Structured matrices such as the Bezoutian, Hankel, Vandermonde and Toeplitz, arise in a wide variety of practical applications. Developing fast algorithms for these matrices is important from view points of practical applications. Though fast and numerically effective algorithms now exist for triangular factorizations and linear systems solutions (see [19,39,76,112]) such algorithms for eigenvalue problems are, however, rare. On the other hand, in many practical instances, all that one needs is the inertia; an explicit knowledge of eigenvalues is not required.

Similar remarks hold for the inertia of large and sparse arbitrary matrices. We have remarked before that it is generally believed that the best way, from numerical view point, to determine the inertia of a matrix, is to explicitly compute the eigenvalues of the matrix. Unfortunately, the methods for eigenvalue computation of large and sparse matrices are not well-developed. Currently, this is an active area of research (see [58,102]).

In view of the above remarks, we may pose the following two problems:

**Problem 3.** Develop fast algorithms for computing the inertia of the Bezoutian, Vandermonde, Hankel, Toeplitz, and other Toeplitz-like matrices, by exploiting the structures of these matrices.
Remark 7.2. We believe that in the development of fast algorithms for a Bezoutian matrix, relation (2.3) might play some role. This is seen as follows.

The inertia of a symmetric matrix can be computed by knowing the signs of its leading principal minors [57, Vol. II]. Again, the leading principal minors of the product of two matrices can be computed in terms of the leading principal minors of the individual matrices using the Cauchy–Binet theorem (see [57, Vol. I]). The leading principal minors of \( U \) in (2.3) are trivially found. All then remains to be done is, to find an \( O(n^2) \) algorithm for computing the leading principal minors of the matrix \( g(A) \) by taking advantage of the simple structure of the companion matrix \( A \). Wilkinson [121] has described a numerically stable scheme for computing the leading principal minors of a matrix using Givens rotations. It is worth-while to find how this numerically stable scheme can be used to compute the leading principal minors of \( g(A) \) in \( O(n^2) \) operations, taking advantage of the companion structure of \( A \).

Problem 4. Develop indirect algorithms (algorithms that do not require knowledge of eigenvalues or solution of any matrix equations) for computing the inertia of a large and sparse matrix.

Remark 7.3. An algorithm, based on the sparse \( LDL^T \) factorization, for large and sparse symmetric matrices exists (see [54]).

Remark 7.4. The inertia method stated in Section 6 is not practical for large and sparse matrices. It is based on the reduction of \( A \) to a Hessenberg matrix, and the Householder and Givens methods for reduction of \( A \) to Hessenberg matrix are well known to destroy the sparsity.

Finally, we state the following problem related to computing the inertia of a nonsymmetric matrix with respect to the unit circle.

Problem 5. Develop a unit-circle analogue of the inertia method of Carlson and Datta described in Section 6.

Acknowledgements

I would like to express my gratitude to Hans Schneider for providing me with some valuable information on the historical developments of some of the inertia theorems reported in this paper and for his remarks on early drafts of this paper.

Thanks are also due to David Carlson for his reading of early drafts, useful remarks, and providing me with an unpublished manuscript of his paper (1998).
I would like to thank my student Daniil Sarkissian for his careful reading and re-reading of various versions of the manuscript and making helpful suggestions and comments.

I would also like to take this opportunity to pay a respectful homage to my Ph.D advisor, the late Professor James L. Howland, for introducing me to the field of inertia and stability, and for his constant encouragement which led to several papers in this area in the early part of my research career.

Finally, thanks are due to Liz Mehren for her help in typing and re-typing various versions of this manuscript with great patience.

References


[111] O. Taussky, Matrices C with $C^n \to 0$. J. Algebra 1 (1964) 5–10.