R. E. KALMAN
Research Institute for Advanced Study,[2] Baltimore, Md.

J. E. BERTRAM
IBM Research Center, Yorktown Heights, N. Y.

Control System Analysis and Design Via the "Second Method" of Lyapunov[1]

I Continuous-Time Systems

The "second method" of Lyapunov is the most general approach currently in the theory of stability of dynamic systems. After a rigorous exposition of the fundamental concepts of this theory, applications are made to (a) stability of linear stationary, linear nonstationary, and nonlinear systems; (b) estimation of transient behavior; (c) control-system optimization; (d) design of relay servos. The discussion is essentially self-contained, with emphasis on the thorough development of the principal ideas and mathematical tools. Only systems governed by differential equations are treated here. Systems governed by difference equations are the subject of a companion paper.

1—Introduction

One of the most important events in the theory of stability of dynamic systems was the publication in 1892 of Lyapunov's famous memoire in a Russian journal. Translated into French in 1907 and reprinted in America in 1947 [1],[3] it is still relatively unknown and unappreciated in the West. By contrast, the so-called "second method" of Lyapunov has achieved virtual pre-eminence in the Soviet Union as the principal mathematical tool in tackling linear and nonlinear stability problems of the most varied type, particularly in the theory of control systems.

The purpose of this paper is to give a self-contained exposition of the second method of Lyapunov, with particular attention to applications in the theory of control. We shall treat here only continuous-time dynamic systems; i.e., those governed by ordinary differential equations. In a companion paper, an analogous treatment is given of discrete-time dynamic systems.

The objective of the so-called "second method" of Lyapunov is this: *To answer questions of stability of differential equations, utilizing the given form of the equations but without explicit knowledge of the solutions.* The name "second method" is an unfortunate but well-entrenched historical misnomer[4]; actually the "second method" is more accurately described as a point of view, a philosophy of approach, rather than a systematic method. At present, much depends on the ingenuity of the user. In the future, we can hope that systematic procedures will be made possible by machine computation. Yet the concept of a Lyapunov function (see below) appears to be very basic; the whole theory of control systems can be studied in a unified way using Lyapunov functions, as we shall show in what follows. Being a relatively new point of view, the "second method" has much promise for the further development of control theory, particularly as regards nonlinear systems.

The principal idea of the second method is contained in the following physical reasoning: *If the rate of change $dE(\mathbf{x})/dt$ of the energy $E(\mathbf{x})$ of an isolated physical system is negative for every possible state $\mathbf{x}$, except for a single equilibrium state $\mathbf{x}_e$, then the energy will continually decrease until it finally assumes its minimum value $E(\mathbf{x}_e)$.* In other words, a dissipative system perturbed from its equilibrium state will always return to it; this is the intuitive concept of stability.

In order to develop this idea into a precise mathematical tool, the foregoing reasoning must be made independent of the physical concept of energy. As a rule there is no natural way of defining energy when the equations of motion are given in a purely mathematical form. The mathematical counterpart of the foregoing reasoning is the following conjectured theorem: *A dynamic system is stable (in the sense that it returns to equilibrium after any perturbation) if and only if there exists a "Lyapunov function," i.e., some scalar function $V(\mathbf{x})$ of the state with the properties: (a) $V(\mathbf{x}) > 0$, $\dot{V}(\mathbf{x}) < 0$ when $\mathbf{x} \neq \mathbf{x}_e$ and (b) $V(\mathbf{x}) = \dot{V}(\mathbf{x}) = 0$ when $\mathbf{x} = \mathbf{x}_e$.* As will be seen (Theorem 1 and 2), this conjecture is correct under slight additional mathematical restrictions. Occasionally, but not necessarily always, it will suffice to take $V = E$.

One of the attractions of the "second method" is its appeal to geometric intuition. This is illustrated by the following two examples.

Example 1. Harmonic Oscillator. Normalizing the angular frequency to 1, a harmonic oscillator is described by the linear differential equation

$$\ddot{x}_1 + x_1 = 0 \qquad (\dot{\ } = d/dt)$$

Letting $\dot{x}_1 = x_2$, this equation can be written in the equivalent form

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 \end{aligned}\right\} \tag{1}$$

The state of the system is represented by the vector $\mathbf{x} = (x_1, x_2)$. The solutions (trajectories) of (1) in the state-plane (phase-plane) are concentric circles about the origin, Fig. 1. The energy of the system is given by

$$E(\mathbf{x}) = E(x_1, x_2) = x_1^2 + x_2^2 = V(\mathbf{x}) \tag{2}$$

The derivative $\dot{V}$ of V along any solution of (1), i.e., along any trajectory in Fig. 1, is given by

---

[1] This research was supported in part by the U. S. Air Force Office of Scientific Research under Contract AF 49(638)-382.

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] The so-called "first method" of Lyapunov deals with stability questions via an explicit representation of the solutions of a differential equation.

Contributed by the Instruments and Regulators Division of The American Society of Mechanical Engineers and presented at the ASME-AIEE-IRE-ISA National Automatic Control Conference, Dallas, Tex., November 4–5, 1959.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 10, 1959. Paper No. 59—NAC-2.

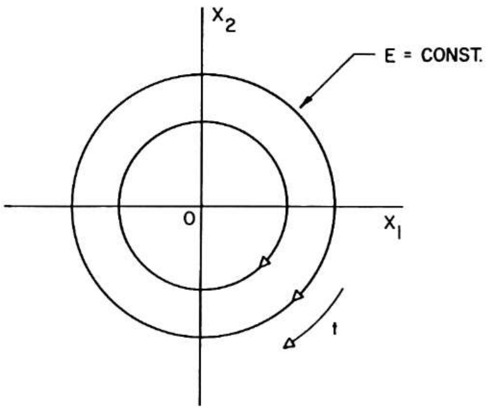


Fig. 1 Phase-plane trajectories of harmonic oscillator

$$\dot{V}(\mathbf{x}) = \frac{dV(\mathbf{x}(t))}{dt} = \dot{\mathbf{x}}'[\text{grad } V(\mathbf{x})]$$

$$= \dot{x}_1 \frac{\partial V}{\partial x_1} + \dot{x}_2 \frac{\partial V}{\partial x_2}$$

$$= 2\dot{x}_1 x_1 + 2\dot{x}_2 x_2$$

which is identically zero by (1). Hence $\dot{V}(\mathbf{x}) = 0$; the energy remains constant—(1) is a conservative system.

Example 2. Asymptotically Stable Nonlinear System. Consider now a slight modification of system (1):

$$\left.\begin{aligned} \dot{x}_1 &= x_2 - ax_1(x_1^2 + x_2^2) \\ \dot{x}_2 &= -x_1 - ax_2(x_1^2 + x_2^2) \end{aligned}\right\} \tag{3}$$

where a is a positive constant.

We again define V by (2), but now of course this quantity is not necessarily the energy of system (3). Calculating as before, the derivative $\dot{V}$ of V along a solution of (3) is

$$\dot{V}(\mathbf{x}) = \frac{dV(\mathbf{x}(t))}{dt} = -2a(x_1^2 + x_2^2)^2 \tag{4}$$

$$= -2aV^2(\mathbf{x})$$

which is negative unless $x_1 = x_2 = 0$. This shows that V is constantly decreasing along any solution of (3), and hence V is a Lyapunov function.

With reference to Fig. 2, it follows also that the trajectories of (3) cross the boundary of every region $V(\mathbf{x}) \leq$ const from the outside toward the inside. From this we are led to the following geometric interpretation of a Lyapunov function:

Let $V(\mathbf{x})$ be a measure of the "distance" of the state $\mathbf{x}$ from the origin in the state space, i.e., $V(\mathbf{x}) > 0$ when $\mathbf{x} \neq \mathbf{0}$ and $V(\mathbf{0}) = 0$. Suppose the distance between the origin and the instantaneous state $\mathbf{x}(t)$ is continually decreasing as $t \to \infty$, i.e., $\dot{V}(\mathbf{x}(t)) < 0$. Then $\mathbf{x}(t) \to \mathbf{0}$.

In Examples 1–2, $V(\mathbf{x})$ is actually the square of the Euclidean distance. In general $V(\mathbf{x}) =$ const turns out to be (because of $\dot{V}(\mathbf{x}) < 0$) a family of concentric closed surfaces surrounding the origin such that the surface $V(\mathbf{x}) = c_1$ lies inside the surface $V(\mathbf{x}) = c_2$ whenever $c_1 < c_2$. Thus we can say that $\mathbf{y}$ is "farther" away from the origin than $\mathbf{x}$ if $V(\mathbf{y}) > V(\mathbf{x})$.

Unlike the energy of a system, the Lyapunov function $V(\mathbf{x})$ is not unique; this is precisely the reason why the "second method" of Lyapunov is a more powerful tool than conventional energy considerations. A system whose energy E decreases *on the average*, but not necessarily at each instant, is stable but E is not a Lyapunov function.

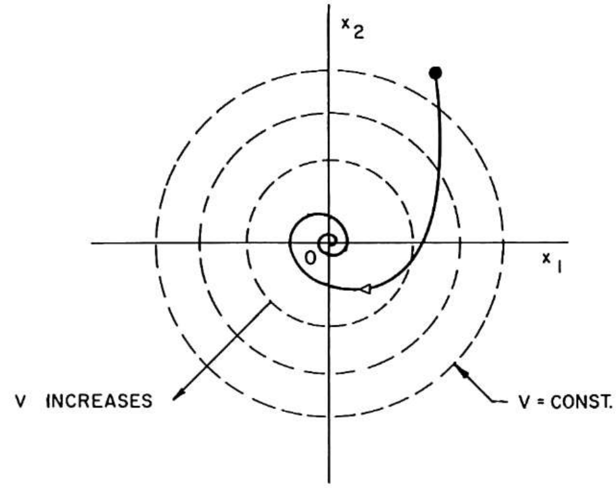


Fig. 2 Phase-plane trajectories of harmonic oscillator with nonlinear damping

2—Outline of Contents

Sections 4–6 together with the Appendix provide the necessary mathematical machinery. Sections 7–10 contain the applications.

Description of Dynamic Systems (Section 4). To apply the "second method," it is necessary to describe dynamic systems from the "state" point of view. Since this may be unfamiliar to many control engineers, an outline of the most important concepts and definitions is provided.

Definitions of Stability (Section 5). In treating general types of systems (nonlinear, nonstationary) it is necessary to make careful distinctions between various intuitive aspects of stability. The necessary definitions are stated precisely and in accordance with current mathematical usage.

The Main Theorems (Section 6). Theorem 1 gives the conditions on a Lyapunov function whose existence guarantees the strongest type of stability. In the corollaries of Theorem 1, these conditions are stated in various weakened forms suitable for certain applications. Theorem 3 summarizes the current state of knowledge concerning the stability of linear, nonstationary systems; Corollaries 3.1–2 provide a purely algebraic procedure for solving many problems in linear stationary systems, which is simpler than the customary methods based on the Laplace transform and is well suited for machine computation. All proofs of this section may be omitted at first reading without loss of continuity.

Applications to Stability Theory (Section 7). An easy consequence of Corollary 3.1 in the linear, stationary case is a rigorous elementary derivation of the Routh-Hurwitz inequalities (Example 6). The paper and pencil derivation is very messy, but the method lends itself well to machine computation. Example 7 gives the construction of a Lyapunov function for the canonic form of feedback systems. Example 8 is an application of the "second method" to an electric circuit consisting of a time-varying resistor, inductor, and capacitor. Example 9 gives a rigorous justification of the usual linearization procedure. Example 10 is a proof of the stability of passive nonlinear networks. Examples 11 and 13 give conditions, similar to the Routh-Hurwitz inequalities, for the stability of some nonlinear systems; Theorem 4 is probably the best available result along these lines which does not depend on the order of the system.

Estimation of Transient Behavior (Section 8). Since a Lyapunov function can be regarded as defining distance from the origin in the state space, its derivative $\dot{V}$ can be used to give a quantitative estimate of the speed with which the origin is approached. This leads to an equivalent time constant even for nonlinear systems (see Example 14). Similarly, the Lyapunov function can be used to estimate the effect of random perturbations and sensitivity to parameter variations, as in Example 15.

Relations With System Optimization (Section 9). If a performance index is defined as the error criterion integrated with respect to time, then the performance index is actually a Lyapunov function—provided the error criterion is not identically zero along any trajectory of the system. This observation (Theorem 5) removes a serious shortcoming of the conventional Wiener theory of quadratic optimization since the latter does not automatically guarantee that the quadratic optimal system is stable. Example 16 shows how a conventional quadratic optimization problem can be solved by means of Corollary 3.1; Example 17 is a rigorous treatment of the same problem using elementary concepts from game theory.

Design of a Relay Servo (Section 10). While exact optimization is usually difficult or impractical, the "second method" can supply some new ideas for approximate optimization. In Example 18 we present a very simple design algorithm of this type for a stable relay servo (again based on Corollary 3.1) which requires only a linear compensating element. The analytical findings are supported by an analog-computer study.

3—Guide to Western Literature

There are very few expositions of the subject in Western languages; most of those available are written from the abstract mathematical point of view. The following comments are intended to facilitate study of the subject from the engineering point of view.

The original work of Lyapunov [1] (in French) is difficult to read because of obsolescent mathematical terminology. An authoritative survey of mathematical results until 1956 is given by Massera [2]. The recent monograph of Hahn [3] (in German) covers existing theoretical results and includes a few applications. There is now an English translation [4] of the well-known book of Malkin. This book, allegedly addressed to engineers, is primarily a detailed and rigorous mathematical treatment of classical problems. There is a section (Chapter VI, par. 3) on the basic theorems of Lyapunov in the well-known book of Lefschetz [5]; see also his paper [6] on recent generalizations. The problem of control-system design has been treated at length by Lur'e [7] (German and English translation) and Letov [8].

Hahn [3] and the survey article of Antosiewicz [9] give many references to Soviet literature. Over the past 10 years, a great deal has been published about the "second method" in the journals *Avtomatika i Telemekhanika* and *Prikladnaya Matematika i Mekhanika* (now both available in English translation).

4—Description of Continuous-Time Dynamic Systems

From now on extensive use will be made of vector notation. Small boldface Roman or Greek letters will denote vectors. Capital boldface Roman or Greek letters will denote matrices. The unit matrix is $\mathbf{I}$, the zero vector or matrix $\mathbf{0}$. The transpose of a vector or matrix is denoted by the prime; the scalar product of $\mathbf{x}$, $\mathbf{y}$ is $\mathbf{x}'\mathbf{y}$, and the quadratic form associated with a square matrix $\mathbf{A}$ is $\mathbf{x}'\mathbf{A}\mathbf{x}$. The norm is denoted by $\|\mathbf{x}\|$; in specific calculations this is to be taken as the Euclidean norm $(\mathbf{x}'\mathbf{x})^{1/2}$. The eigenvalues of a matrix $\mathbf{F}$ are $\lambda_i(\mathbf{F})$.

In this paper we shall study systems governed by the *vector differential equation:*

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{u}(t), t), \qquad -\infty < t < +\infty \tag{5}$$

This is equivalent to the set of n scalar differential equations

$$dx_i/dt = f_i(x_1, \ldots, x_n, u_1(t), \ldots, u_m(t), t) \tag{5$_i$}$$
$$(i = 1, \ldots, n)$$

The vector $\mathbf{x}$ is the *state* of the system (5), its components x_i are the *state variables*. The vector $\mathbf{u}(t)$ is the *control* function (or *forcing function* or *input*) of the system (5), its components $u_i(t)$ are the *control variables*. The system is specified by the vector-valued function $\mathbf{f}$. The integer n is the *order* of the system.

If $\mathbf{u}(t) \equiv \mathbf{0}$ for all t, we say that (5) is *free* (*unforced*):

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t) \tag{5-F}$$

It is always assumed that the function $\mathbf{f}$ in (5) is sufficiently smooth so that the equation has a unique solution starting at any initial state $\mathbf{x}_0$ at any time t_0.

To make this idea precise, we first fix $\mathbf{u}(t)$ in (5), so that we might as well consider (5-F). We assume that *there exists a unique vector function $\boldsymbol{\phi}(t; \mathbf{x}_0, t_0)$, differentiable in t, such that, for any fixed $\mathbf{x}_0$, t_0*

(i) $\boldsymbol{\phi}(t_0; \mathbf{x}_0, t_0) = \mathbf{x}_0$
(ii) $d\boldsymbol{\phi}(t; \mathbf{x}_0, t_0)/dt = \mathbf{f}(\boldsymbol{\phi}(t; \mathbf{x}_0, t_0), t)$

in some interval $|t - t_0| \leq a(t_0)$. The function $\boldsymbol{\phi}$ is called a *solution* of the system (5-F) or (5).

In older texts on differential equations, particularly in the engineering literature, it is customary to write $\mathbf{x}(t)$ for the solution. This "informal" usage will be employed here only rarely; reasons of clarity demand in most cases (as in Massera [2] and Hahn [3]) the use of the explicit notation just introduced. Thus $\boldsymbol{\phi}(t; \mathbf{x}_0, t_0)$ means: a solution of (5), with fixed $\mathbf{u}(t)$, going through state $\mathbf{x}_0$ at time t_0, observed at time t. We also call $\boldsymbol{\phi}$ the *transition function* since it shows how $\mathbf{x}(t_0)$ is transformed into $\mathbf{x}(t)$. With this interpretation, $\mathbf{f}$ in (5) may be called the *infinitesimal transition function* since the differential equation specifies the infinitesimal state transition $\mathbf{x} \rightarrow \mathbf{x} + d\mathbf{x}$ corresponding to the infinitesimal change in time $t \rightarrow t + dt$.

Now if the differential equation (5) is to represent a physical phenomenon, one would always assume also that *$\boldsymbol{\phi}(t; \mathbf{x}_0, t_0)$ is continuous in all of its arguments* (of course, continuity in t is already demanded by (ii) above!). A sufficient condition for the validity of the foregoing assumptions is provided by the following classical theorem (Ref. [11], Chap. 1, Theorems 2.3 and 7.1):

EXISTENCE, UNIQUENESS, AND CONTINUITY THEOREM. *Let $\mathbf{f}(\mathbf{x}, t)$ be continuous in $\mathbf{x}$, t, and satisfy a Lipschitz condition in some region about any $\mathbf{x}_0$, t_0:*

$$R(\mathbf{x}_0, t_0) = \begin{cases} \|\mathbf{x} - \mathbf{x}_0\| \leq b(\mathbf{x}_0) \\ |t - t_0| \leq c(t_0) \end{cases} \quad (b, c > 0)$$

i.e., the following condition is satisfied for $(\mathbf{x}, t)$, $(\mathbf{y}, t)$ in $R(\mathbf{x}_0, t_0)$

$$\|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}(\mathbf{y}, t)\| \leq k\|\mathbf{x} - \mathbf{y}\|$$

where k is a positive constant which depends only on b, c. THEN

1 *There exists a unique solution $\boldsymbol{\phi}(t; \mathbf{x}_0, t_0)$ of* (5-F) *starting at* $\mathbf{x}_0$, t_0, for all $|t - t_0| \leq a(t_0)$,

$$a(t_0) \geq \text{Min}\{c(t_0), b(\mathbf{x}_0)/M(\mathbf{x}_0, t_0)\}$$

where $M(\mathbf{x}_0, t_0)$ is the maximum assumed by the continuous function $\|\mathbf{f}(\mathbf{x}, t)\|$ in the closed, bounded set $R(\mathbf{x}_0, t_0)$.

2 *In some small neighborhood of $\mathbf{x}_0$, t_0 the solution is a continuous function of its arguments.*

Note that the Lipschitz condition implies continuity of $\mathbf{f}$ in $\mathbf{x}$, but not necessarily in t; it is implied by bounded partial derivatives in $\mathbf{x}$. It should be borne in mind that the *local* Lipschitz condition required by the theorem only implies the desired properties of a solution *near* $\mathbf{x}_0$, t_0. From it we cannot conclude the existence of solutions for arbitrarily large values of t as the following example shows:

Example 3. Finite Escape Time. Consider the first-order nonlinear differential equation

$$\dot{x}_1 = x_1^2$$

The requirements of the existence and uniqueness theorem are obviously satisfied.

By separation of variables, the solution is seen to be

$$-\frac{1}{\phi_1(t;\ x_{01}, t_0)} + \frac{1}{x_{01}} = t - t_0$$

Now if $x_{01} > 0$, the solution will increase with t; therefore $\phi_1(t;\ x_{01}, t_0)$ tends to infinity as t tends to $t_1 = t_0 + 1/x_{01}$. As t increases beyond t_1, the original solution which started at t_0, $x_{01} > 0$ cannot be continued, except possibly by letting the state $x_1(t_1)$ "jump" from $+\infty$ to $-\infty$, which is clearly not the way physical systems behave.

Following Markus [17], one says in this case that the solution of the differential equation escapes to infinity in finite time or has *finite escape time*. (This can be always avoided by replacing t by $t' = t/(t_\infty - t)$.) In order that a differential equation represent a physical system, this *possibility must be ruled out by explicit assumption to the contrary.*

If the Lipschitz condition holds for $\mathbf{f}$ *everywhere*, there can be no finite escape time. To see this (it does not follow directly from the theorem!), let

$$\|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}(\mathbf{y}, t)\| \leqq k\|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}$, $\mathbf{y}$, t.

Integrating both sides of (5-F) and using this, and (8-A)

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| \leqq \|\mathbf{x}_0\| + \left\|\int_{t_0}^{t} \mathbf{f}(\boldsymbol{\phi}(\tau;\ \mathbf{x}_0, t_0), \tau)d\tau\right\|$$

$$\leqq \|\mathbf{x}_0\| + k\int_{t_0}^{t} \|\boldsymbol{\phi}(\tau;\ \mathbf{x}_0, t_0)\|d\tau$$

This is easily seen to imply, by the well-known Gronwall-Bellman lemma (Reference [49], p. 35), that

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| \leqq [\exp k(t - t_0)]\|\mathbf{x}_0\| \qquad (6)$$

which is less than infinity for any finite $(t - t_0)$.

However, for some purposes the requirement of a *global* Lipschitz condition is too restrictive; for instance, in case of servos with ramp inputs.

If the foregoing hypotheses concerning $\boldsymbol{\phi}$ are satisfied, the differential equation (5) is said to represent a *continuous-time dynamic system.* We then call $\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)$ a *motion* (or *trajectory* or *transition function*) of the dynamic system. Frequently, it is convenient to define a dynamic system axiomatically by means of its motions satisfying the foregoing hypotheses [10]. These axioms may be stated as follows:

(A_1) (i) $\boldsymbol{\phi}(t_0;\ \mathbf{x}_0, t_0) = \mathbf{x}_0$ for all $\mathbf{x}_0, t_0$ (7)

(ii) $\boldsymbol{\phi}(t_2;\ \boldsymbol{\phi}(t_1;\ \mathbf{x}_0, t_0), t_1)$
$= \boldsymbol{\phi}(t_2;\ \mathbf{x}_0, t_0)$ for all $\mathbf{x}_0, t_0, t_1, t_2$ (8)

(*existence* and *uniqueness;* see Fig. 3.)

(A_2) *$\boldsymbol{\phi}$ is continuous with respect to all arguments.*

(A_3) *$\boldsymbol{\phi}$ is defined for all* $\mathbf{x}_0, t_0, t$ (no finite escape time).

A state $\mathbf{x}_e$ of a free dynamic system (5-F) is an *equilibrium state* if

$$\mathbf{f}(\mathbf{x}_e;\ t) = \mathbf{0} \quad \text{for all} \quad t$$

or, equivalently,

$$\boldsymbol{\phi}(t;\ \mathbf{x}_e, 0) = \mathbf{x}_e \quad \text{for all} \quad t$$

In other words, a motion passing through an equilibrium state at any time is actually at the same state at all times.

A dynamic system (5) is *stationary* if

$$\mathbf{f}(\mathbf{x}, \mathbf{u}(t), t) \equiv \mathbf{f}(\mathbf{x}, \mathbf{u}(t)) \qquad \text{for all} \qquad \mathbf{x}, \mathbf{u}(t)$$

i.e., if $\mathbf{f}$ does not depend explicitly on time (but may depend implicitly on time through $\mathbf{u}(t)$). For a free system, this is equivalent to saying that every motion is *invariant under translation in time,* i.e., that

$$\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) = \boldsymbol{\phi}(t + \tau;\ \mathbf{x}_0, t_0 + \tau)$$

for any $\mathbf{x}_0$, t_0, τ. A system which is both free and stationary is said to be *autonomous.*

A dynamic system (5) is *linear* if $\mathbf{f}$ is a linear function of $\mathbf{x}$ and $\mathbf{u}$. Assuming for simplicity that $\mathbf{0}$ is an equilibrium state, i.e., that

$$\mathbf{f}(\mathbf{0}, \mathbf{0}, t) = \mathbf{0} \qquad (8\text{-A})$$

it follows that a linear system (5) can be written in the form

$$d\mathbf{x}/dt = \mathbf{F}(t)\mathbf{x} + \mathbf{D}(t)\mathbf{u}(t) \qquad (5\text{-L})$$

where

$$\mathbf{F}(t) = [\partial f_i(t)/\partial x_j]$$

$$\mathbf{D}(t) = [\partial f_i(t)/\partial u_j]$$

are $n \times n$ resp. $n \times m$ matrices depending on time. In a stationary linear system $\mathbf{F}$ and $\mathbf{D}$ are constant matrices.

For $t \geqq t_0$, the general solution of (5-L) can be written as [11]

$$\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) = \boldsymbol{\Phi}(t, t_0)\mathbf{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\,\mathbf{D}(\tau)\mathbf{u}(\tau)d\tau \qquad (9)$$

where $\boldsymbol{\Phi}(t, t_0)$ is called the *transition matrix* of the system (5-L). Its elements may be regarded as the impulse responses under appropriate excitations and observations [12, 13]. Thus $\phi_{ij}(t, t_0)$ is the response of the ith state variable to an excitation defined by (taking $\mathbf{D}(t) \equiv \mathbf{I}$):

$$u_j(t) = \delta(t - t_0);\ u_k(t) = 0 \quad \text{when} \quad k \neq j$$

with $\mathbf{x}_0 = \mathbf{0}$.

As a consequence of linearity, relations (7) and (8) now imply (set $\mathbf{u}(t) \equiv \mathbf{0}$ in (9)!)

$$\boldsymbol{\Phi}(t_0, t_0) = \mathbf{I} \qquad \text{for all } t_0 \qquad (7\text{-L})$$

and

$$\boldsymbol{\Phi}(t_2, t_0) = \boldsymbol{\Phi}(t_2, t_1)\boldsymbol{\Phi}(t_1, t_0) \quad \text{for all} \quad t_0, t_1, t_2 \qquad (8\text{-L})$$

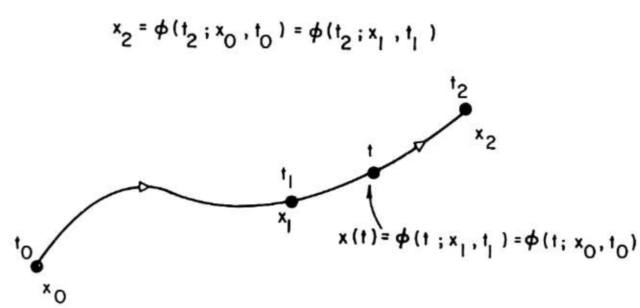


Fig. 3 Notation for motions of a dynamic system

From this, it follows at once also that

$$\Phi^{-1}(t, t_0) = \Phi(t_0, t) \qquad \text{for all } t, t_0 \tag{9}$$

Another useful property of the transition matrix, an immediate consequence of its definition by (9), is that it satisfies its own differential equation:

$$d\Phi(t, t_0)/dt = \mathsf{F}(t)\Phi(t, t_0) \quad \text{for all } t, t_0 \tag{10}$$[5]

In a stationary linear system $\Phi(t, t_0)$ depends only on the difference $t - t_0$. In fact, it is easy to show [5, 11] that

$$\Phi(t, t_0) = \sum_{k=0}^{\infty} [(t - t_0)\mathsf{F}]^k/k! = \exp(t - t_0)\mathsf{F} \tag{11}$$

As usually stated, the "second method" of Lyapunov applies only to free dynamic systems, with equilibrium state at $\mathbf{0}$; that is to say, to deviations about some fixed motion. To see this, let $\mathbf{y}(t)$ be some fixed motion of (5), corresponding to a fixed choice of $\mathbf{u}(t)$. Let

$$\mathbf{x} = \mathbf{y}(t) + \mathbf{z}$$

Then

$$\begin{aligned} d\mathbf{x}/dt &= d\mathbf{y}(t)/dt + d\mathbf{z}/dt \\ &= \mathbf{f}(\mathbf{y}(t) + \mathbf{z}, \mathbf{u}(t), t) \\ &= \mathbf{f}(\mathbf{y}(t), \mathbf{u}(t), t) + \mathbf{g}(\mathbf{z}, t) \end{aligned}$$

Since $\mathbf{y}(t)$ is a solution of (5), it follows that

$$d\mathbf{z}/dt = \mathbf{g}(\mathbf{z}, t),$$

$$\mathbf{g}(\mathbf{0}, t) = \mathbf{0} \quad \text{for all } t$$

so that, for fixed $\mathbf{u}(t)$, deviations from the motion $\mathbf{y}(t)$ are a free dynamic system.

The description of dynamic systems via the concept of "state" (instead of more restricted concepts like impulse response, frequency response, transfer function, etc., which are invalid for nonlinear and sometimes even for nonstationary linear systems) is essential for the understanding of the "second method." In fact, this is necessary even for a precise definition of stability. Though not yet in common use, the "state" method has been applied recently to numerous control problems; see references [12–16] which also contain much pedagogical material in regard to setting up of equations, interpretation of solutions in geometric terms, etc. The idea of "state" is well known, of course; it has been used in mathematics more or less explicitly at least since Poincaré.

5—Concepts of Stability

On closer examination, the intuitive idea of stability turns out to be very complex. This is particularly true in nonlinear dynamic systems. As a result, a large variety of precise definitions of stability have been proposed, often differing in subtle ways which are confusing to the uninitiated. Here we shall present the principal definitions used at present and discuss their relationships to one another.

The question of stability concerns deviation about some *fixed* motion. By the remarks at the end of the previous section, it follows that it suffices to consider deviations from an equilibrium state $\mathbf{x}_e$ of a free dynamic system.

Perhaps the oldest concept of stability is this: If the system (5-F) is perturbed slightly from its equilibrium state at the origin, all subsequent motions remain in a correspondingly small neighborhood of the origin. For instance, harmonic oscillators (Example 1) possess this type of stability. The precise definition goes back to Lyapunov:

(S_1) An equilibrium state $\mathbf{x}_e$ of a free dynamic system is *stable* if for every real number $\epsilon > 0$ there exists a real number $\delta(\epsilon, t_0) > 0$ such that $\|\mathbf{x}_0 - \mathbf{x}_e\| \leq \delta$ implies

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e\| \leq \epsilon \qquad \text{for all } t \geq t_0$$

Fig. 4 illustrates the definition. Note that stability in the sense of Lyapunov (definition (S_1)) is a *local* concept; it refers to behavior near the equilibrium state since one does not know *a priori* how small δ in the definition may have to be chosen.

Observe further that *if there is stability for some initial time t_0 there is stability for any other initial time t_1, provided all motions are continuous in the initial state* (Axiom (A_2)).

In fact, if $t_1 < t_0$ define (possible, since $\boldsymbol{\phi}$ is continuous in t!)

$$\Omega(\mathbf{x}_1, t_1, t_0) = \operatorname*{Max}_{t_1 \leq t \leq t_0} \|\boldsymbol{\phi}(t;\ \mathbf{x}_1, t_1) - \mathbf{x}_e\| \tag{12}$$

This function is clearly 0 when $\mathbf{x}_1 = \mathbf{x}_e$; elementary but somewhat tedious arguments show that, due to (A_2), Ω is continuous in $\mathbf{x}_1$. Hence there is some $\nu(\delta(\epsilon, t_0), t_1) > 0$ such that $\|\mathbf{x}_1 - \mathbf{x}_e\| \leq \nu$ implies $\|\boldsymbol{\phi}(t;\ \mathbf{x}_1, t_0) - \mathbf{x}_e\| \leq \delta$ in the interval $t_1 \leq t \leq t_0$. By stability at t_0, it follows that $\|\boldsymbol{\phi}(t;\ \mathbf{x}_1, t_1) - \mathbf{x}_e\| \leq \epsilon$ for all $t \geq t_1$. We can take $\delta(\epsilon, t_1) = \nu$. If $t_0 \leq t_1$, merely let $\Omega = \|\boldsymbol{\phi}(t_0;\ \mathbf{x}_1, t_1)\|$.

In most engineering applications, one is interested in a stronger kind of stability; namely, the motion should return to equilibrium after any small perturbation. The classical definition of Lyapunov is the following:

(S_2) An equilibrium state $\mathbf{x}_e$ of a free dynamic system is *asymptotically stable* if
(i) it is stable and
(ii) every motion starting sufficiently near $\mathbf{x}_e$ converges to $\mathbf{x}_e$ as $t \rightarrow \infty$. In other words, there is some real constant $r(t_0) > 0$ and to every real number $\mu > 0$ there corresponds a real number $T(\mu, \mathbf{x}_0, t_0)$ such that $\|\mathbf{x}_0 - \mathbf{x}_e\| \leq r(t_0)$ implies

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e\| \leq \mu \text{ for all } t \geq t_0 + T$$

See Fig. 5. Asymptotic stability is also a local concept; one does not know *a priori* how small $r(t_0)$ may have to be. If a free *linear* system is asymptotically stable, then $\|\mathbf{x}_1 - \mathbf{x}_e\| < \|\mathbf{x}_0 - \mathbf{x}_e\|$ implies that $T(\mu, \mathbf{x}_1, t_0) < T(\mu, \mathbf{x}_0, t_0)$ for any μ and t_0, since the initial state is factored out in $\boldsymbol{\phi}$ (see (9)). In other words, among motions starting at the same *distance* from $\mathbf{x}_e$ none will remain at a distance larger than μ from $\mathbf{x}_e$ at arbitrarily large values of time. This is expressed by the following definition due to Massera [18]:

(S_3) An equilibrium state $\mathbf{x}_e$ of a free dynamic system is *equiasymptotically stable* if
(i) it is stable and
(ii) every motion starting sufficiently near $\mathbf{x}_e$ converges to $\mathbf{x}_e$ as $t \rightarrow \infty$ uniformly in $\mathbf{x}_0$. In other words, T in definition (S_2-ii) is of the form $T(\mu, r(t_0), t_0)$.

A somewhat puzzling aesthetic deficiency of definitions (S_2 — S_3) is that one has to assume *explicitly* that (S_1) holds. The following example shows that (S_2-ii) does not imply (S_1) or (S_3-ii) when $n > 1$.

Example. Consider the second-order system in polar co-ordinates $\mathbf{x} = (r, \theta)$, $0 \leq r < \infty$, $0 \leq \theta < 2\pi$:

$$\left.\begin{aligned} \dot{r} &= [\dot{g}(\theta, t)/g(\theta, t)]r \\ \dot{\theta} &= 0 \end{aligned}\right\}$$

[5] In the literature [5, 11], any nonsingular matrix function of time satisfying (10) is called a *fundamental matrix* of (5-L). In our case, the transition matrix is a fundamental matrix made *unique* through the requirement (7-L).

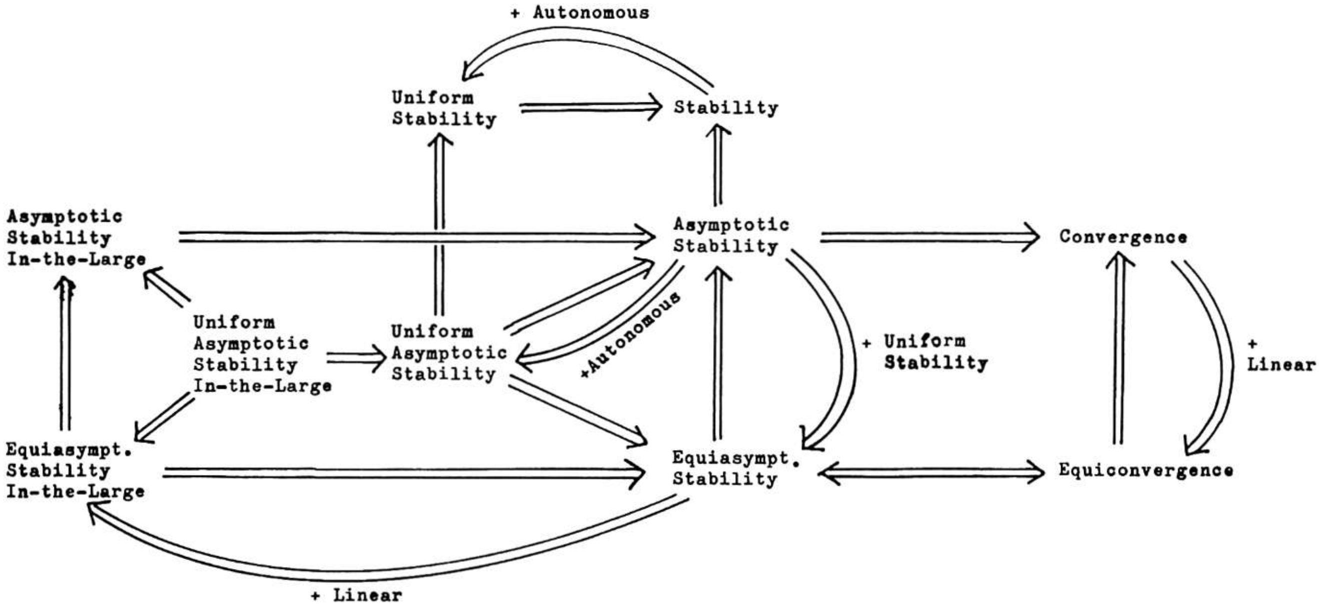


Fig. 6 Interrelations between stability concepts

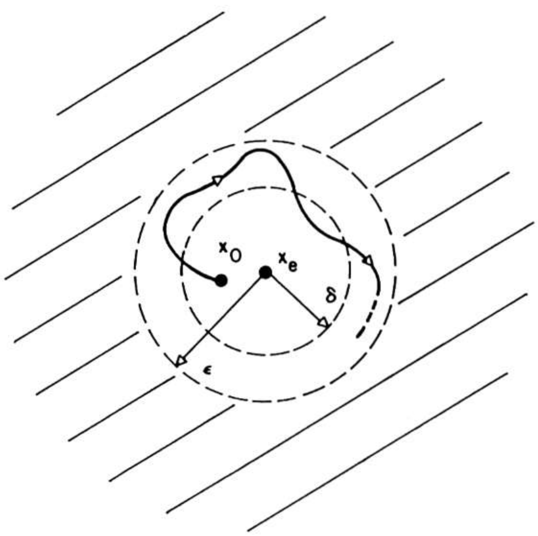


Fig. 4 Definition of stability

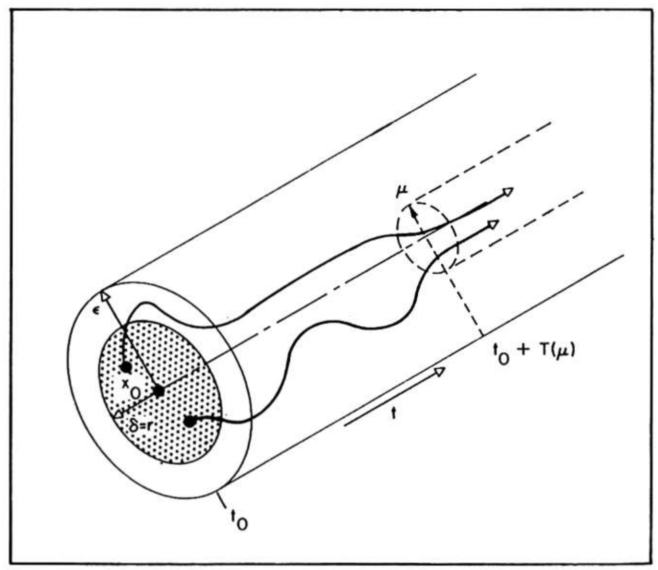


Fig. 5 Definition of asymptotic stability

where

$$g(\theta, t) = \frac{\sin^2 \theta}{\sin^4 \theta + (1 - t \sin^2 \theta)^2} + \frac{1}{1 + t^2}$$

Then

$$r(t;\ r_0, \theta_0, t_0) = [g(\theta_0, t)/g(\theta_0, t_0)]r_0$$

$$\theta(t;\ r_0, \theta_0, t_0) = \theta_0$$

The motion is continuous with respect to r_0, θ_0. (S_2-ii) is satisfied. However, (S_1), (S_3-ii) do not hold because at $t_1 = \sin^{-2}\theta_0$, $g(\theta_0, t_1) > \sin^{-2}\theta_0$, which can be made arbitrarily large as $\theta_0 \rightarrow \pm\pi$. Here the function $T(\mu, \mathbf{x}_0, t_0)$ is not continuous in $\mathbf{x}_0$ along the ray $\theta_0 = \pm\pi$.

On the other hand, (S_3-ii) *implies* (S_1) *if all motions are continuous in* $\mathbf{x}_0$.

In fact, let $t_1 = t_0 + T(\epsilon, r(t_0), t_0)$ and define

$$\Lambda(\mathbf{x}_0, t_0, T) = \underset{t_0 \leqq t \leqq t_1}{\mathrm{Max}} \left\| \boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e \right\|$$

Since T does *not* depend on $\mathbf{x}_0$ by (S_3-ii), it follows by (A_2), as in (12), that Λ is continuous in $\mathbf{x}_0$. Hence there is some $\delta(\epsilon, t_0)$ such that $\|\mathbf{x}_0 - \mathbf{x}_e\| \leqq \delta$ implies

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e\| \leqq \epsilon \text{ for all } t_0 \leqq t \leqq t_0 + T$$

By (S_3-ii), the inequality then holds for all $t \geqq t_0$.

(S_3) obviously implies (S_2). (S_2) *is equivalent to* (S_3) *if the dynamic system is of the first order* [18]. In fact, if $x_e < y < x$ (scalars), then for all t, t_0 $\phi(t;\ y, t_0) < \phi(t; x, t_0)$, for otherwise by continuity in t one would have $z = \phi(t_1;\ x, t_0) = \phi(t_1;\ y, t_0)$ which says that the motion starting at z, t_1 may be either at x or y at t_0, contradicting axiom (A_1). Therefore, if

$$|x| \leqq r(t_0),\ T(\mu, x_0, t_0) \leqq \mathrm{Max}\ T(\mu, \pm r(t_0), t_0)$$

Another important property of linear systems is that stability is independent of the distance of the initial state from $\mathbf{x}_e$. This is expressed by

(S_4) an equilibrium state $\mathbf{x}_e$ of a free dynamic system is *asymptotically* [(S_5): *equiasymptotically*] *stable in the large* if

(i) it is stable and

(ii) every motion converges to $\mathbf{x}_e$ [converges to $\mathbf{x}_e$ uniformly in $\mathbf{x}_0$ for $\|\mathbf{x}_0\| \leqq r$ where r is fixed but arbitrarily large] as $t \to \infty$.

If a system is linear (S_2-S_5) are equivalent.

A motion is said to be *bounded* for every $\mathbf{x}_0$, t_0 if there is some constant $B(\mathbf{x}_0, t_0)$ such that $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| \leqq B$ for all $t \geqq t_0$. The motion is *equibounded* if $B(\mathbf{x}_0, t_0) \leqq B(r, t_0)$ for all $\|\mathbf{x}_0\| \leqq r$. These properties are implied by asymptotic resp. equiasymptotic stability in the large as is seen from the proof that (S_3-ii) implies (S_1).

If a system is autonomous (= free + stationary), then δ, r, T in the preceding definitions do not depend on t_0. It is not hard to show [2] that the same holds also if the function $\mathbf{f}(\mathbf{x}, t)$ is periodic in t. There are systems which have this property without being autonomous or periodic. This motivates the various definitions of *uniform* (in time) stability.

(S_6) *Uniform stability* is stability such that δ does not depend on t_0.

It turns out that *asymptotic and equiasymptotic stability are equivalent if they are uniform in* t_0.

In fact, take any $\mu > 0$. Since the foregoing hypotheses imply (S_6), there is some corresponding $\delta(\mu) > 0$. Consider any $\mathbf{x}$ in the region $\|\mathbf{x} - \mathbf{x}_e\| \leqq r$. By uniform (in t_0) convergence, find a $T(\delta/2, \mathbf{x})$ such that

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}, t_0) - \mathbf{x}_e\| \leqq \delta/2 \quad \text{for all } t \geqq t_0 + T$$

By continuity with respect to the initial state, there is some neighborhood $U(\mathbf{x})$ of $\mathbf{x}$ such that if $\mathbf{y}$ is in $U(\mathbf{x})$ then

$$\|\boldsymbol{\phi}(t_0 + T;\ \mathbf{y}, t_0) - \mathbf{x}_e\| \leqq \delta$$

so that, because of (S_6), the motion starting at $\mathbf{y}$ remains within distance of at most μ of $\mathbf{x}_e$ for all $t \geqq t_0 + T$.

Since $\|\mathbf{x} - \mathbf{x}_e\| \leqq r$ is compact, it is covered by definition by a finite number of $U(\mathbf{x})$, say, $U(\mathbf{x}_1), \ldots, U(\mathbf{x}_N)$. Hence define

$$T(\mu, r) = \operatorname*{Max}_i \{T(\delta/2, \mathbf{x}_i)\} < \infty$$

which completes the proof.

The foregoing argument leads to a finite value of $T(\mu, r, t_0)$ if r and T are dependent on t_0. Then we can conclude: (S_2-ii) *together with* (S_6) *implies* (S_3).

These observations motivate the following definition, which is standard:

(S_7) *Uniform asymptotic stability* is equiasymptotic stability such that δ, r, T do not depend on t_0.

It is clear that uniform asymptotic stability involves the explicit assumption of uniform stability. Unlike in the case of nonuniform stability, where (S_3-ii) implies (S_1), this assumption cannot be dispensed with as the following example due to Massera shows:

Example 4. Impulse Response With Growing Peaks. Consider the first-order linear system

$$\dot{x}_1 = (4t \sin t - 2t)x_1$$

By separation of variables and integrating, the impulse response of this system (= 1×1 transition matrix) is

$$\phi_{11}(t, t_0) = \exp\,(4 \sin t - 4t \cos t - t^2 - 4 \sin t_0 + 4t_0 \cos t_0 + t_0^2)$$

which is less than

$$\exp\,[4(2 + |t| + |t_0|) - t^2 + t_0^2]$$

which shows that $\|\boldsymbol{\phi}\| \to 0$ with $t \to \infty$ uniformly in t_0 (and of course uniformly in $\|\mathbf{x}_0\| \leqq r$, by linearity.) However,

$$\phi_{11}((2n + 1)\pi, 2n\pi) = \exp\,[\pi(4 - \pi)(4n + 1)]$$

The "peaks" of the impulse response increase indefinitely as $t_0 \to \infty$. There is no uniform stability; in fact, the motions are not even uniformly bounded.

This example motivates also the following definition:

(S_8) An equilibrium state $\mathbf{x}_e$ of a free dynamic system is *uniformly asymptotically stable in the large* if

(i) it is uniformly stable,

(ii) it is uniformly bounded, i.e., given any $r > 0$ there is some $B(r)$ such that $\|\mathbf{x}_0 - \mathbf{x}_e\| \leqq r$ implies $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e\| \leqq B$ for all $t \geqq t_0$,

(iii) every motion converges to $\mathbf{x}_e$ as $t \to \infty$ uniformly in t_0 and $\|\mathbf{x}_0\| \leqq r$, when r is fixed but arbitrarily large; i.e., given any $r > 0$ and $\mu > 0$ there is some $T(\mu, r)$ such that $\|\mathbf{x}_0 - \mathbf{x}_1\| \leqq r$ implies $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0) - \mathbf{x}_e\| \leqq \mu$ for all $t \geqq 0\ t_0 + T$.

One can define a type of boundedness by taking $B(\mathbf{x}_0)$; however, this condition, with (S_8-iii), would imply (S_8-ii), by arguments similar to those concerning uniform asymptotic stability.

By a slight extension of Example 4, it follows that (S_8-iii) does not imply (S_8) in general. A case where this is true occurs when $\mathbf{f}(\mathbf{x}, t)$ satisfies a uniform Lipschitz condition, i.e.,

$$\|\mathbf{f}(\mathbf{x}, t) - \mathbf{f}(\mathbf{y}, t)\| \leqq k\|\mathbf{x} - \mathbf{y}\|$$

where k is a constant independent of $\mathbf{x}$, $\mathbf{y}$, or t. To prove this assertion, one merely makes use of the relation (6).

The different types of stability, summarizing the discussion of this section, are displayed in Fig. 6.

It is apparent that concepts of stability are closely related to concepts of convergence. When there are many of the latter, there are correspondingly many types of stability. For instance, in the study of stochastic systems, one can talk about stability in probability, stability in mean, almost sure stability, etc. It is not hard in principle to apply the "second method" to such problems as has been done by Bertram and Sarachik [19].

6—The Main Theorems

The application of the "second method" to stability problems consists of defining a Lyapunov function with appropriate properties whose existence implies the desired type of stability. We shall consider uniform asymptotic stability in the large and define the class of Lyapunov functions for this case. By weakening the various requirements on Lyapunov functions, we obtain other stability theorems as a by-product. See also Massera [2], Hahn [3], and Malkin [4].

Theorem 1. (Lyapunov) *Consider the continuous-time, free dynamic system*

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t) \tag{5-F}$$

where $\mathbf{f}(\mathbf{0}, t) = \mathbf{0}$ *for all* t.

SUPPOSE *there exists a scalar function* $V(\mathbf{x}, t)$ *with continuous first partial derivatives with respect to* $\mathbf{x}$ *and* t *such that* $V(\mathbf{0}, t) = 0$ *and*

(i) $V(\mathbf{x}, t)$ *is positive definite; i.e., there exists a continuous, nondecreasing scalar function* α *such that* $\alpha(0) = 0$ *and, for all* t *and all* $\mathbf{x} \neq \mathbf{0}$

$$0 < \alpha(\|\mathbf{x}\|) \leqq V(\mathbf{x}, t)$$

(see Fig. 7);

(ii) *There exists a continuous scalar function* γ *such that* $\gamma(0) = 0$ *and the derivative* $\dot{V}$ *of* V *along the motion starting at* t, $\mathbf{x}$ *satisfies, for all* t *and all* $\mathbf{x} \neq \mathbf{0}$,

$$
\left.\begin{aligned}
\dot{V}(\mathbf{x}, t) &\equiv \frac{dV(\mathbf{x}, t)}{dt}\Bigg|_{\substack{\text{along motion}\\ \text{starting at } t,\, \mathbf{x}}} \\
&= \frac{dV(\boldsymbol{\phi}(\tau;\ \mathbf{x}_0, t), \tau)}{d\tau}\Bigg|_{\tau = t} \\
&= \lim_{h \to 0}\ [V(\mathbf{x} + h\mathbf{f}(\mathbf{x}, t), t + h) - V(\mathbf{x}, t)]/h \\
&= \partial V/\partial t + (\operatorname{grad} V)'\mathbf{f}(\mathbf{x}, t) \\
&\leqq -\gamma(\|\mathbf{x}\|) < 0^{6};
\end{aligned}\right\} \tag{13}
$$

(iii) *There exists a continuous, nondecreasing scalar function such that* $\beta(0) = 0$ *and, for all* t,

$$V(\mathbf{x}, t) \leqq \beta(\|\mathbf{x}\|)^{7};$$

(iv) $\alpha(\|\mathbf{x}\|) \to \infty$ with $\|\mathbf{x}\| \to \infty$

THEN *the equilibrium state* $\mathbf{x}_e = \mathbf{0}$ *is uniformly asymptotically stable in the large;* $V(\mathbf{x}, t)$ *is called a Lyapunov function of the system* (5-F).

Corollary 1.1. *The following conditions are sufficient for the various weaker types of stability:*

(*a*) *Uniform asymptotic stability:* (i–iii).

(*b*) *Equiasymptotic stability in the large:* (i–ii), (iv).

(*c*) *Equiasymptotic stability:* (i–ii).

(*d*) *Uniform stability:* (i), (iii), *and* (ii_1): $\dot{V}(\mathbf{x}, t) \leqq 0$ *for all* $\mathbf{x}, t$.

(*e*) *Stability:* (i–ii_1).

(*f*) *No finite escape time:* (i), (iv), *and* (ii_2): $\dot{V}(\mathbf{x}, t) \leqq cV(\mathbf{x}, t)$ *for all* $\mathbf{x}, t$; *c being a positive constant.*

Corollary 1.2. *For a continuous-time autonomous dynamic system*

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}) \tag{5-A}$$

$\mathbf{f}(\mathbf{0}) = \mathbf{0}$, *equiasymptotic stability in the large is assured by the existence of scalar function* $V(\mathbf{x})$ *with continuous first partial derivatives with respect to* $\mathbf{x}$, *such that* $V(\mathbf{0}) = 0$ *and*

(i-A) $V(\mathbf{x}) > 0$ *for all* $\mathbf{x} \neq \mathbf{0}$

(ii-A) $\dot{V}(\mathbf{x}) < 0$ *for all* $\mathbf{x} \neq \mathbf{0}$

(iv-A) $V(\mathbf{x}) \to \infty$ *with* $\|\mathbf{x}\| \to \infty$

Corollary 1.3. *In Corollary 1.2, Condition (ii-A) may be replaced by:*

(ii_1-A) $\dot{V}(\mathbf{x}) \leqq 0$ *for all* $\mathbf{x}$

(ii_2-A) $\dot{V}(\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0))$ *does not vanish identically in* $t \geqq t_0$ *for any* t_0 *and any* $\mathbf{x}_0 \neq \mathbf{0}$.

Proof of Theorem 1. Using (ii) we see that V is decreasing along any motion:

$$V(\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0), t) - V(\mathbf{x}_0, t_0) = \int_{t_0}^{t} \dot{V}(\boldsymbol{\phi}(\tau;\ \mathbf{x}_0, t_0), \tau)d\tau < 0, \qquad t > t_0 \tag{14}$$

(*a*) To prove uniform stability, consider any $\epsilon > 0$. Take $\delta(\epsilon) > 0$ such that $\beta(\delta) < \alpha(\epsilon)$—this is possible because β is continuous and 0 at 0; see Fig. 7. Then if $\|\mathbf{x}_0\| \leqq \delta$, t_0 being arbitrary, we have using (ii) and (14) (see Fig. 8), for all $t \geqq t_0$

$$
\begin{aligned}
\alpha(\epsilon) > \beta(\delta) &\geqq V(\mathbf{x}_0, t_0) \\
&\geqq V(\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0), t) \\
&\geqq \alpha(\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\|)
\end{aligned}
$$

But since α is nondecreasing and positive, this implies

$$\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| < \epsilon; \ t \geqq t_0, \quad \|\mathbf{x}_0\| \leqq \delta$$

for arbitrary t_0, which is what was to be proved. Note that (iv) is not needed and that it is sufficient to have $\dot{V} \leqq 0$.

(*b*) Now we prove that $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| \to 0$ with $t \to \infty$ uniformly in t_0 and $\|\mathbf{x}_0\| \leqq r$.

Take any positive constant c_1 and find an $r > 0$ satisfying $\beta(r) < \alpha(c_1)$. Take any initial state $\|\mathbf{x}_0\| \leqq r$. By part (*a*) of the proof, then $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| \leqq c_1$ for all $t \geqq t_0$, t_0 being arbitrary.

Now take any $0 < \mu \leqq \|\mathbf{x}_0\|$. Find a $\nu(\mu) > 0$ such that $\beta(\nu) < \alpha(\mu)$. Denote by $c_2(\mu, r) > 0$ the minimum of the continuous function $\gamma(\|\mathbf{x}\|)$ on the compact set $\nu(\mu) \leqq \|\mathbf{x}\| \leqq c_1(r)$. Define $T(\mu, r) = \beta(r)/c_2(\mu, r) > 0$.

Suppose now that $\|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| > \nu$ over the interval $t_0 \leqq t \leqq t_1 = t_0 + T$. Again by (ii) and (14), we have

$$
\begin{aligned}
0 < \alpha(\nu) &\leqq V(\boldsymbol{\phi}(t_1;\ \mathbf{x}_0, t_0), t_1) \\
&\leqq V(\mathbf{x}_0, t_0) - (t_1 - t_0)c_2 \\
&\leqq \beta(r) - Tc_2 = 0
\end{aligned}
$$

which is a contradiction. Hence for some t in the interval t_0, t_1, say t_2, we have $\|\mathbf{x}_2\| = \|\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)\| = \nu$. Therefore

$$
\begin{aligned}
\alpha(\|\boldsymbol{\phi}(t;\ \mathbf{x}_2, t_2)\|) &\leqq V(\boldsymbol{\phi}(t;\ \mathbf{x}_2, t_2), t) \\
&\leqq V(\mathbf{x}_2, t_2) \\
&\leqq \beta(\nu) < \alpha(\mu)
\end{aligned}
$$

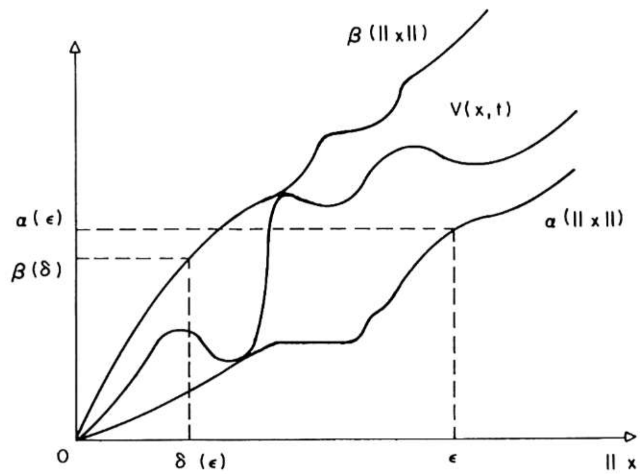
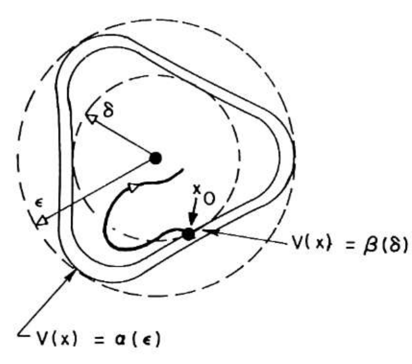


Fig. 7 Definition of V, α, β



$\alpha(\|\mathbf{x}(t)\|) \leqq V(\mathbf{x}(t)) \leqq V(\mathbf{x}_0) \leqq \beta(\delta) < \alpha(\epsilon)$

Fig. 8 Proof of asymptotic stability

---

[6] This is slightly weaker than the usual requirement that $\dot{V}$ be negative definite, since we do not need the hypothesis that γ is nondecreasing.

[7] In the literature, particularly the Russian, this condition is sometimes called: "V has an infinitely small upper bound."

for all $t \geqq t_2$. Hence

$$\|\phi(t;\ x_0, t_0)\| < \mu \quad \text{for all } t \geqq t_0 + T(\mu, r) \geqq t_2$$

which proves uniform asymptotic stability.

(*c*) To prove uniform asymptotic stability in the large, i.e., that the constant r can be chosen to be arbitrarily large, observe that by (iv) there exists for any r a constant $c_1(r)$ such that $\beta(r) < \alpha(c_1)$.

Moreover, uniform boundedness is automatic, on setting $B(r) = c_1(r)$. Q.E.D.

Proof of Corollary 1.1. (*a*) This is contained in the proof of Theorem 1.

(*b*) Let $\beta^*(\|\mathbf{x}\|, t)$ be the maximum of $V(\mathbf{y}, t)$ for $\|\mathbf{y}\| \leqq \|\mathbf{x}\|$. Replacing $\beta(\|\mathbf{x}\|)$ by $\beta^*(\|\mathbf{x}\|, t)$, the proof of Theorem 1 proceeds as before, but now δ and T are also functions of t_0.

(*c*) Use β^* instead of β and omit Part (*c*) of the proof of Theorem 1.

(*d*) Part (*a*) of the proof of Theorem 1.

(*e*) Use β^* defined previously and proceed as in Part (*a*) of the proof of Theorem 1. Then δ is also a function of t_0.

(*f*) An easy calculation shows that (ii$_2$) implies

$$V(\phi(t;\ x_0, t_0)) \leqq e^{c(t-t_0)}V(x_0, t_0)$$

and by (i)

$$\alpha(\|\phi(t;\ x_0, t_0)\|) \leqq e^{c(t-t_0)}V(x_0, t_0) < \infty$$

Because of (iv), $\|\mathbf{x}\| = \infty$ implies $\alpha(\|\mathbf{x}\|) = \infty$ which is impossible for any finite t. Q.E.D.

Proof of Corollary 1.2. By (iv), given any $a > 0$ there is a $b(a) > 0$ such that $V(\mathbf{x}) > a$ when $\|\mathbf{x}\| > b$. (If (iv-A) is missing and only local stability is of interest, pick $b > 0$ arbitrarily.) Let

$$c(\|\mathbf{x}\|) = \{\min_{\mathbf{y}} V(\mathbf{y});\ \|\mathbf{y}\| = \|\mathbf{x}\|\}$$

and then define

$$\alpha(\|\mathbf{x}\|) = \min_{\mathbf{y}} \{V(\mathbf{y});\ \|\mathbf{x}\| \leqq \|\mathbf{y}\| \leqq b(c(\|\mathbf{x}\|))\}$$

$$\beta(\|\mathbf{x}\|) = \max_{\mathbf{y}} \{V(\mathbf{y});\ \|\mathbf{y}\| \leqq \|\mathbf{x}\|\}$$

$$\gamma(\|\mathbf{x}\|) = \min_{\mathbf{y}} \{-\dot{V}(\mathbf{y});\ \|\mathbf{y}\| = \|\mathbf{x}\|\}$$

The functions α, β, γ obviously satisfy the hypotheses of Theorem 1. Q.E.D.

Proof of Corollary 1.3. It suffices to show that (ii$_1$-A), (ii$_2$-A), together with (i-A), (iv-A) imply asymptotic stability in the large for stationary systems. By (iv-A) and (ii$_1$-A), for every $\mathbf{x}_0$ there is a constant $c_1(\|\mathbf{x}_0\|)$ such that $\|\phi(t;\ \mathbf{x}_0, t_0)\| \leqq c_1$ for all $t \geqq t_0$, t_0 being arbitrary since the system is stationary.

Suppose $\|\phi(t;\ \mathbf{x}_0, t_0)\| \geqq \nu > 0$. By (i-A) and (ii$_1$-A) the integral

$$-V(\mathbf{x}_0) < \int_{t_0}^{t} \dot{V}(\phi(\tau;\ \mathbf{x}_0, t_0))d\tau \leqq 0$$

is nonincreasing and bounded as $t \to \infty$, which implies that $\dot{V}$ is identically zero on the positive limit set [11] of $\mathbf{x}_0$, which is contained in the region $\nu \leqq \|\mathbf{x}\| \leqq c_1$. Since the limit set is itself a motion, this contradicts (ii$_2$-A). Hence every motion will eventually enter the set $\|\mathbf{x}\| \leqq \nu$. But if ν is chosen so that $\beta(\nu) < \alpha(\mu)$, then asymptotic stability in the large follows, as in the proof of Theorem 1. Q.E.D.

As already pointed out, there is no explicit prescription in general for finding Lyapunov functions satisfying the requirements of Theorem 1 or of its corollaries. For this reason, it is very important to know when a Lyapunov function with certain properties exists, so that the search for one should not be in vain! It turns out that this question is closely related to the concept of the uniform asymptotic stability. We now quote what is at present the best available result along these lines [2]:

Theorem 2. (Massera) *Let the function* $\mathbf{f}$ *defining the differential equation* (5-F) *be Lipschitzian. Assume also that* $\mathbf{f}(\mathbf{0}, t) = \mathbf{0}$ *and that the equilibrium state* $\mathbf{x}_e = \mathbf{0}$ *is uniformly asymptotically stable in the large.*

Then there exists a Lyapunov function $V(\mathbf{x}, t)$ *which is infinitely differentiable with respect to* $\mathbf{x}$, t, *and satisfies all of the hypotheses of Theorem 1.*

Thus the existence of a Lyapunov function as required by Theorem 1 is necessary and sufficient for uniform asymptotic stability in the large.

PROOF: See Reference [2], Theorem 23*.

The stability problem in nonstationary linear systems is one of considerable subtlety. Roughly speaking, the requirement of "uniformity" must be present in the formal definitions of stability, in order that it should coincide with the intuitive concept of stability. The troubles are illustrated by the following example:

Example 5. Stable System With Unbounded Step Response. Consider the nonstationary linear system

$$\dot{x}_1 + x_1/t = y_1(t)$$

for values of $0 < t_0 \leqq t$. (For values of t_0 below 0, the equation has finite escape time (see Example 3), and therefore does not define a dynamic system.) The impulse response (which in this case is also the 1×1 transition matrix) of this system is

$$\phi_{11}(t, t_0) = t_0/t;\ t \geqq t_0 > 0$$

The system is evidently asymptotically stable, but not uniformly asymptotically stable: If $t_0 > 0$ and $T = (t - t_0)$ is a fixed constant then

$$\phi_{11}(t, t_0) = t_0/(T + t_0) > 1/2$$

for t_0 sufficiently large, no matter how large T is.

Now suppose that the initial state at some $t_0 > 0$ is zero. Let the input $y_1(t)$ be a unit step function occurring at $t = t_0$. Then the unit step response is given by

$$x_1(t) = \int_{t_0}^{t} \phi_{11}(t, \tau)d\tau$$

$$= (t - t_0^2/t)/2 \qquad (t \geqq t_0 > 0)$$

which tends to infinity with t!

The difficulty here is due to the fact that the impulse response does not tend to 0 exponentially as in an asymptotically stable linear *stationary* system. It so happens that, in the linear case, exponential convergence to equilibrium is equivalent to uniform asymptotic stability. The latter, in turn, is equivalent to the existence of a Lyapunov function or the fact that every bounded excitation produces a bounded response. These facts are summarized by:

Theorem 3. (Lyapunov-Perron-Malkin) *Consider a continuous-time, linear dynamic system*

$$d\mathbf{x}/dt = \mathbf{F}(t)\mathbf{x} + \mathbf{D}(t)\mathbf{u}(t) \tag{5-L}$$

subject to the restrictions

(i) $\|\mathbf{F}(t)\| \leqq c_1{}^8 < \infty$ *for all* t

(ii) $0 < c_2 \leqq \|\mathbf{D}(t)\mathbf{x}\| \leqq c_3 < \infty$ *for all* $\|\mathbf{x}\| = 1$, *all* t

THEN *the following propositions concerning this system are equivalent:*

(A) *Any uniformly bounded excitation*

$$\|\mathbf{u}(t)\| \leqq c_4 < \infty \quad (t \geqq t_0)$$

[8] In this theorem and its proof, $c_1, c_2, \ldots$ are used to denote various fixed positive constants. We continue to use this convention in the sequel.

gives rise to a uniformly bounded response for all $t \geqq t_0$

$$\|\mathbf{x}(t)\| = \left\|\boldsymbol{\phi}(t, t_0)\mathbf{x}_0 + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\mathbf{D}(\tau)\mathbf{u}(\tau)d\tau\right\|$$

$$\leqq c_5(c_4, \|\mathbf{x}_0\|) < \infty; \tag{14-A}$$

(*B*) *For all $t \geqq t_0$,* $\displaystyle\int_{t_0}^{t} \|\boldsymbol{\Phi}(t, \tau)\|d\tau \leqq c_6 < \infty$;

(*C*) *The equilibrium state $\mathbf{x}_e = \mathbf{0}$ of the free system is uniformly asymptotically stable;*

(*D*) *There are positive constants c_7, c_8 such that, whenever $t \geqq t_0$,*

$$\|\boldsymbol{\Phi}(t, t_0)\| \leqq c_7 e^{-c_8(t-t_0)}; \tag{15}$$

(*E*) *Given any positive-definite matrix $\mathbf{Q}(t)$ continuous in t and satisfying for all $t \geqq t_0$*

$$0 < c_9\mathbf{I} \leqq \mathbf{Q}(t) \leqq c_{10}\mathbf{I} < \infty^{9} \tag{16}$$

the scalar function defined by

$$V(\mathbf{x}, t) = \int_{t}^{\infty} \|\boldsymbol{\Phi}(\tau, t)\mathbf{x}\|^2_{\mathbf{Q}(\tau)}d\tau^{10} \tag{17}$$

$$= \|\mathbf{x}\|^2_{\mathbf{P}(t)} \tag{18}$$

exists[11] *and is a Lyapunov function for the free system satisfying the requirements of Theorem 1, with its derivative along the free motion starting at* $\mathbf{x}$, t *being*

$$\dot{V}(\mathbf{x}, t) = -\|\mathbf{x}\|^2_{\mathbf{Q}(t)}$$

Propositions (*A*)-(*D*) relate to different abstract notions of stability. The equivalence of (*A*) and (*B*) was first proved by Perron [20]. Zadeh used (*A*) as his basic definition of stability [21]. By proposition (*E*), the preceding notions of stability are equivalent to the existence of a Lyapunov function. Malkin [22] was the first to prove that (*C*) is equivalent to (*E*); his method of proof implied condition (*B*), hence, by Perron's result, the proof of the entire theorem. See also Antosiewicz and Davis [23]. The proof given below, based on unpublished work of Kalman, is much simpler than the original proof of Malkin and Perron, due to the systematic use of the transition matrix.

None of the propositions in Theorem 3 offers a method of determining whether a given system is uniformly asymptotically stable or not without computing its transition matrix for all t, t_0. Such a method is known so far only in the case of linear stationary systems and is already contained in Lyapunov's original memoire:

Corollary 3.1. (Lyapunov) *The equilibrium state $\mathbf{x}_e = \mathbf{0}$ of a continuous-time, free, linear, stationary dynamic system*

$$d\mathbf{x}/dt = \mathbf{F}\mathbf{x} \tag{5-FLS}$$

is asymptotically stable (*a*) *if and* (*b*) *only if given any symmetric, positive-definite matrix $\mathbf{Q}$ there exists a symmetric, positive-definite matrix $\mathbf{P}$ which is the unique solution of the set of $n(n+1)/2$*[12] *linear equations*

$$\mathbf{F}'\mathbf{P} + \mathbf{P}\mathbf{F} = -\mathbf{Q} \tag{19}$$

and $\|\mathbf{x}\|^2_{\mathbf{P}}$ is a Lyapunov function for (5-FLS).

A slight extension of this result is useful for some purposes:

Corollary 3.2. (Kalman) *The real parts of the eigenvalues of a* (*constant*) *matrix $\mathbf{F}$ are $< \sigma$ if and only if given any symmetric, positive-definite matrix $\mathbf{Q}$ there exists a symmetric, positive-definite matrix $\mathbf{P}$ which is the unique solution of the set of $n(n+1)/2$ linear equations.*

$$-2\sigma\mathbf{P} + \mathbf{F}'\mathbf{P} + \mathbf{P}\mathbf{F} = -\mathbf{Q} \tag{20}$$

Proof of Theorem 3. (Kalman) The proof consists mainly of using the multiplicative property (8-L) of transition matrices and various technicalities involving norms and positive-definite matrices.

To show that (*A*) implies (*B*). If (*B*) did not hold, for at least one pair (i^*, j^*) of subscripts the integral

$$\int_{t_0}^{t} |\phi_{i^*j^*}(t, \tau)|d\tau$$

must tend to ∞ as $(t - t_0) \to \infty$. By (ii), $1/\epsilon_3 \leqq \|\mathbf{D}^{-1}(\tau)\| \leqq 1/c_2$ for all τ. Therefore define, for all $t_0 \leqq \tau \leqq t$

$$u_k(\tau) = [\mathbf{D}^{-1}(\tau)]_{kj^*} \operatorname{sgn} \phi_{i^*j^*}(t, \tau)$$

where

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \tag{21}$$

Then $u(\tau)$ is bounded; we can take $c_4 = 1/c_2$. Using this and (75), with $\mathbf{x}_0 = \mathbf{0}$,

$$\infty > c_5(c_4, 0) \geqq \|\mathbf{x}(t)\|$$

$$\geqq n^{-1/2} \sum_{i=1}^{n} |x_i(t)|$$

$$\geqq \left|\sum_{k=1}^{n} \int_{t_0}^{t} \phi_{i^*j^*}(t, \tau)d_{j^*k}(\tau)u_k(\tau)d\tau\right|$$

$$= \int_{t_0}^{t} |\phi_{i^*j^*}(t, \tau)|d\tau$$

By assumption, the last integral tends to infinity with $(t - t_0)$, which is a contradiction.

To show that (*B*) implies (*C*). (This is the only nontrivial part of the proof.) First we show that (*B*) implies

$$\|\boldsymbol{\Phi}(t, t_0)\| \leqq c_{11} < \infty \qquad (t \geqq t_0) \tag{22}$$

By (i) and (*B*) we can write

$$\infty > c_1c_6 \geqq \int_{t_0}^{t} \|\boldsymbol{\Phi}(t, \tau)\| \cdot \|\mathbf{F}(\tau)\|d\tau$$

$$\geqq \left\|\int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\mathbf{F}(\tau)d\tau\right\|$$

Since the transition matrix satisfies its own differential equation (10), we obtain, after using repeatedly (8-L)

$$-\frac{d}{d\tau}\boldsymbol{\Phi}^{-1}(\tau, t) = \boldsymbol{\Phi}^{-1}(\tau, t)\mathbf{F}(\tau) = \boldsymbol{\Phi}(t, \tau)\mathbf{F}(\tau)$$

Therefore

$$\infty > c_1c_6 \geqq \left\|\int_{t_0}^{t} \left[\frac{d\boldsymbol{\Phi}(t, \tau)}{d\tau}\right] d\tau\right\|$$

$$= \|\boldsymbol{\Phi}(t, t_0) - \mathbf{I}\|, \qquad t \geqq t_0$$

which implies (22).

Using (*B*), (22), and (8-L),

$$\infty > c_6c_{11} \geqq \int_{t_0}^{t} \|\boldsymbol{\Phi}(t, \tau)\| \cdot \|\boldsymbol{\Phi}(\tau, t_0)\|d\tau$$

$$\geqq \|\boldsymbol{\Phi}(t, t_0)\|(t - t_0), \qquad t \geqq t_0$$

[9] If $\mathbf{A}$, $\mathbf{B}$ are two matrices, we write $\mathbf{A} < \mathbf{B}$ [$\mathbf{A} \leqq \mathbf{B}$] to express the fact that $\mathbf{B} - \mathbf{A}$ is positive-definite [semidefinite].

[10] We write (see Appendix) $\mathbf{x}'\mathbf{A}\mathbf{x} = \|\mathbf{x}\|^2_{\mathbf{A}}$ where $\mathbf{A}$ is a symmetric, positive-definite matrix.

[11] I.e., the integral defined by (17) is finite for all finite values of $\mathbf{x}$ and t.

[12] Since $\mathbf{P}$ is symmetric, it has precisely this many unknown elements in (19).

Letting

$$T(\mu, r) = \frac{c_6 c_{11}}{\mu} r,$$

it follows that

$$\|\Phi(t, t_0)\mathbf{x}_0\| \leqq \mu$$

when $\|\mathbf{x}_0\| \leqq r$ and $t \geqq T + t_0$, so that we have uniform asymptotic stability.

To show that (C) implies (D). Since uniform asymptotic stability implies uniform stability (see definitions (S_7), (S_6)), the transition matrix is uniformly bounded. In fact, take any $\epsilon > 0$, $0 < \|\mathbf{x}_0\| \leqq \delta(\epsilon)$. By linearity:

$$\delta^{-1}\|\phi(t;\ \mathbf{x}_0, t_0)\| \leqq \|\Phi(t, t_0)\| \cdot \delta^{-1}\|\mathbf{x}_0\| \leqq \|\Phi(t, t_0)\| \leqq \delta^{-1}\epsilon = c_7/2$$

By uniform asymptotic stability, let $T = T(2^{-1}, 1)$ so that

$$\|\Phi(t_0 + T, t_0)\| \leqq 1/2$$

independently of t_0. By induction and (77)

$$\|\Phi(t_0 + kT, t_0)\| \leqq \|\Phi(t_0 + kT, t_0 + (k-1)T)\| \ldots \|\Phi(t_0 + T, t_0)\| \leqq 2^{-k}$$

Hence if $c_8 = (\log 2)/T > 0$, then

$$\|\Phi(t, t_0)\| \leqq c_7 e^{-c_8(t - t_0)} \qquad (t \geqq t_0) \tag{15}$$

To show that (D) implies (A). By (15), (ii), and (78)

$$\left\|\phi(t, t_0)\mathbf{x}_0 + \int_{t_0}^{t} \Phi(t, \tau)\mathbf{D}(\tau)\mathbf{u}(\tau)d\tau\right\| \leqq c_{11}\|\mathbf{x}_0\| + c_3c_4c_7 \int_{t_0}^{t} e^{-c_8(t-\tau)}\, d\tau \leqq c_{11}\|\mathbf{x}_0\| + c_3c_4c_7c_8^{-1} < \infty$$

so that (14-A) holds.

By Theorem 1, (E) implies (C).

To show that (D) implies (E). This is essentially Theorem 2; we give an independent proof. Proceeding as above and using (16)

$$V(\mathbf{x}, t) \leqq c_{10} \int_{t}^{\infty} \|\Phi(\tau, t)\mathbf{x}\|^2 d\tau$$

By (15)

$$\leqq \frac{c_7^2 c_{10}}{2c_8}\|\mathbf{x}\|^2 = \beta(\|\mathbf{x}\|) < \infty \tag{23}$$

Further, using (i), (16), and (10)

$$V(\mathbf{x}, t) \geqq c_1^{-2}c_9 \int_{t}^{\infty} \|\mathbf{F}(\tau)\|^2\|\Phi(\tau, t)\mathbf{x}\|^2 d\tau$$

$$\geqq c_1^{-2}c_9\|\mathbf{x}\|^2 = \alpha(\|\mathbf{x}\|) > 0, \qquad \mathbf{x} \neq \mathbf{0}$$

The derivative $\dot{V}$ of the Lyapunov function V along the free motion starting at t, $\mathbf{x}$ is obtained by direct computation

$$\left.\begin{aligned} \dot{V}(\mathbf{x}, t) &= \lim_{h \to 0} \frac{1}{h}\,[V(\phi(t + h;\ \mathbf{x}, t),\ t + h) - V(\mathbf{x}, t)] \\ &= \lim_{h \to 0} -\frac{1}{h}\left[\int_{t}^{t+h} \|\Phi(\tau, t)\mathbf{x}\|^2_{\mathbf{Q}(\tau)}d\tau\right] \\ &= -\|\mathbf{x}\|^2_{\mathbf{Q}(t)} \leqq -c_9\|\mathbf{x}\|^2 = -\gamma(\|\mathbf{x}\|) < 0 \quad \text{when } \mathbf{x} \neq \mathbf{0} \end{aligned}\right\} \tag{24}$$

This completes the proof of Theorem 3.

Proof of Corollary 3.1. (a) Let $\mathbf{P} > \mathbf{0}$ satisfy (19). Define

$$V(\mathbf{x}) = \|\mathbf{x}\|^2_{\mathbf{P}} = \mathbf{x}'\mathbf{P}\mathbf{x} > 0 \qquad \text{when } \mathbf{x} \neq \mathbf{0}$$

as a tentative Lyapunov function for system (5-FLS). Then, by (13)

$$\dot{V}(\mathbf{x}) = (\text{grad } V)'\mathbf{F}\mathbf{x} = 2(\mathbf{P}\mathbf{x})'\mathbf{F}\mathbf{x} = \mathbf{x}'(\mathbf{F}'\mathbf{P} + \mathbf{P}\mathbf{F})\mathbf{x} \tag{25}$$

and by (19)

$$\dot{V}(\mathbf{x}) = -\mathbf{x}'\mathbf{Q}\mathbf{x} = -\|\mathbf{x}\|^2_{\mathbf{Q}} < 0 \text{ when } \mathbf{x} \neq \mathbf{0}$$

so that V satisfies the hypotheses of Corollary 1.2; therefore (5-FLS) is asymptotically stable.

(b) Suppose (5-FLS) is asymptotically stable. Since it is stationary, it is then also uniformly asymptotically stable. Therefore, by Theorem 3 (C) and (E), the matrix $\mathbf{P}(t)$ defined by (18) exists. Assuming $\mathbf{Q}(t) = \mathbf{Q} =$ const it follows then by the stationarity of (5-FLS) that $\mathbf{P}(t) = \mathbf{P} =$ const. Also, the derivative of the Lyapunov function $V(\mathbf{x})$ defined by (18) is $-\mathbf{x}'\mathbf{Q}\mathbf{x}$, which is equal to (25), so that $\mathbf{P}$ satisfies (19).

To show that $\mathbf{P}$ defined by (18) is the unique solution of (19) let $\bar{\mathbf{P}}$ be any solution of (19). Then, by the definition (11) of the transition matrix in the stationary case

$$\mathbf{P} = \int_0^\infty (\exp\ \mathbf{F}'\tau)\mathbf{Q}(\exp\ \mathbf{F}\tau)d\tau$$

$$= -\int_0^\infty (\exp\ \mathbf{F}'\tau)(\mathbf{F}'\bar{\mathbf{P}} + \bar{\mathbf{P}}\mathbf{F})(\exp\ \mathbf{F}\tau)d\tau$$

Since $\mathbf{F}$ commutes with $\exp\ \mathbf{F}\tau$ for all τ

$$\mathbf{P} = -\int_0^\infty \frac{d}{d\tau}\,[(\exp\ \mathbf{F}'\tau)\bar{\mathbf{P}}(\exp\ \mathbf{F}\tau)]d\tau$$

$$= -(\exp\ \mathbf{F}'\tau)\bar{\mathbf{P}}\,(\exp\ \mathbf{F}\tau)\Big|_0^\infty = \bar{\mathbf{P}}$$

the last equality following from asymptotic stability. Q.E.D.

Incidentally, Bellman [24] proves that (19) has a unique (but not necessarily positive definite) solution for any $\mathbf{Q}$ whenever $\lambda_i(\mathbf{F}) + \lambda_j(\mathbf{F}) \neq 0$ for all i, j.

Proof of Corollary 3.2. The assumptions on $\mathbf{F}$ require that

$$Re\lambda_i(\mathbf{F}) < \sigma$$

From the characteristic equation of a matrix it follows that this is equivalent to:

$$Re\lambda_i(\mathbf{F} - \sigma\mathbf{I}) < 0$$

Replacing $\mathbf{F}$ in (19) by $\mathbf{F} - \sigma\mathbf{I}$ yields (20). (Notice that this argument is equivalent to the "shift" theorem in the theory of the Laplace transform.) Q.E.D.

7—Applications to Stability Theory

7.1 Routh-Hurwitz Conditions for Linear Stationary Systems. Analogously to the various methods associated with the names of Routh and Hurwitz, Corollary 3.1 provides a purely algebraic procedure [25] for testing whether a given free, linear, stationary system is asymptotically stable:

(a) Take an arbitrary symmetric, positive-definite matrix $\mathbf{Q}$ (say, the unit matrix);

(b) Solve (19) for the unknown components of the symmetric matrix $\mathbf{P}$. This is a set of $n(n+1)/2$ linear equations in the unknown elements $p_{11}, \ldots, p_{1n};\ p_{22}, \ldots, p_{2n};\ \ldots;\ p_{nn}$ of $\mathbf{P}$. (The requirement that this set of equations have a nonzero determinant is implied by the Routh-Hurwitz inequalities.)

(c) Check to see if $\mathbf{P}$ is positive-definite. It suffices to check if all leading principal minors[13] of $\mathbf{P}$ are positive. This supplies the Routh-Hurwitz inequalities.

[13] I.e., the determinants of the following submatrices of $\mathbf{P}$: (1) $\mathbf{P}$ itself, (2) $\mathbf{P}$ with the nth row and column deleted, . . ., (n) the 1-1 matrix consisting of p_{11}.

This procedure can be used to derive the Routh-Hurwitz inequalities as in Example 6. However, this is mainly of theoretical interest since the derivations are very complicated when the matrix $\mathbf{F}$ contains more than about 4 or 5 nonzero coefficients.

The principal significance of this procedure is that it provides a convenient and efficient algorithm to determine stability by machine computation. This is far from trivial in the case of large-scale systems since the Routh-Hurwitz conditions in their usual form apply only to a single, nth-order equation (which has in general n different coefficients). Before the Routh-Hurwitz conditions can be applied to a general linear system of n first-order equations (which in general has n^2 different coefficients), it is necessary to compute the characteristic equation of the matrix $\mathbf{F}$; this is far from a trivial job when the number of nonzero coefficients in $\mathbf{F}$ is large. These difficulties are avoided when (19) is used. Another efficient computation procedure for determining stability of general linear stationary systems, due to Schwarz [26], reduces $\mathbf{F}$ to a *canonic form* by means of elementary matrix transformations. This is also related to the "second method" as Example 7 shows.

The concept of a Lyapunov function as in Corollary 3.1 provides a convenient link between the Routh-Hurwitz conditions and other aspects of control-system theory, such as the integral of the quadratic error. This connection, which has come to surface only recently [27], answers an oft-heard but superficial dictum in elementary courses or textbooks, that "the Routh-Hurwitz criteria do not provide a *quantitative* measure of stability." Much further research is needed to completely elucidate these connections.

Example 6. Routh-Hurwitz Conditions for General Second-Order Case. Let $\mathbf{Q} = \mathbf{I}$ and write out explicitly (19) in the case when $\mathbf{F}$ is a 2×2 matrix:

$$\begin{bmatrix} f_{11} & f_{21} \\ f_{12} & f_{22} \end{bmatrix} \cdot \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} + \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} \cdot \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

This system of equations is equivalent to

$$\begin{bmatrix} 2f_{11} & 2f_{21} & 0 \\ f_{12} & f_{11} + f_{22} & f_{21} \\ 0 & 2f_{12} & 2f_{22} \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{22} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

which has a unique solution if and only if the determinant of the matrix $\mathfrak{F}$ on the left is nonzero:

$$\begin{aligned} \det \mathfrak{F} &= 4(f_{11} + f_{22})(f_{11}f_{22} - f_{12}f_{21}) \\ &= n^2(\mathrm{tr}\ \mathbf{F})(\det\ \mathbf{F}) \neq 0 \end{aligned} \qquad (26)$$

Assuming (26), after tedious elementary computations we find

$$\mathbf{P} = \frac{-1}{2(\mathrm{tr}\ \mathbf{F})(\det\ \mathbf{F})} \begin{bmatrix} \det\ \mathbf{F} + f_{21}^2 + f_{22}^2 & -(f_{12}f_{22} + f_{21}f_{11}) \\ -(f_{12}f_{22} + f_{21}f_{11}) & \det\ \mathbf{F} + f_{11}^2 + f_{12}^2 \end{bmatrix} \qquad (27)$$

This matrix is positive definite if and only if

$$\det \mathbf{P} = \frac{(f_{11} + f_{22})^2 + (f_{12} - f_{21})^2}{2(\mathrm{tr}\ \mathbf{F})^2(\det\ \mathbf{F})} > 0 \qquad (28)$$

$$p_{11} = -\frac{\det\ \mathbf{F} + f_{21}^2 + f_{22}^2}{2(\mathrm{tr}\ \mathbf{F})(\det\ \mathbf{F})} > 0 \qquad (29)$$

But (28) implies that

$$\det\ \mathbf{F} = f_{11}f_{22} - f_{12}f_{21} > 0 \qquad (30)$$

and this fact, with (29), implies that

$$\mathrm{tr}\ \mathbf{F} = f_{11} + f_{22} < 0 \qquad (31)$$

and (30–31) together imply (26).

To put these results in a better-known form, let

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix} \qquad (32)$$

Then (30) and (31) require that $a_1 > 0$, $a_2 > 0$ which are the well-known Routh-Hurwitz inequalities for a constant-coefficient linear differential equation of second order.

Example 7. Canonic Form for Linear Stationary Systems. Consider the system shown in Fig. 9. The infinitesimal state-transition matrix $\mathbf{F}$ of this system is:

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ -a_1 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -a_2 & 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & 0 & \ldots & -a_{n-1} & -a_n \end{bmatrix} \qquad (33)$$

In a little-known paper, Schwarz [26] has proved the following: *If a real constant matrix $\mathbf{F}$ is similar to a matrix of form (33) (i.e., $\mathbf{F} = \mathbf{T}^{-1}\mathbf{C}\mathbf{T}$) then the number of eigenvalues $\lambda_i(\mathbf{F})$ of $\mathbf{F}$ which have negative real parts is equal to the number of positive terms in the sequence*

$$a_n, \quad a_n a_{n-1}, \quad \ldots, \quad a_n a_{n-1} \ldots a_1$$

provided none of the a_i is zero.

Under mild restrictions Schwarz [26] also gives an algorithm for computing $\mathbf{C}(\mathbf{F})$ by means of elementary transformations.

When all the a_i are positive, it is easy to show that the system

$$d\mathbf{x}/dt = \mathbf{C}\mathbf{x} \qquad (34)$$

is asymptotically stable. (The matrix (32) is a special case of this!) In fact, let $\mathbf{P}$ be the matrix whose elements on the main diagonal are given

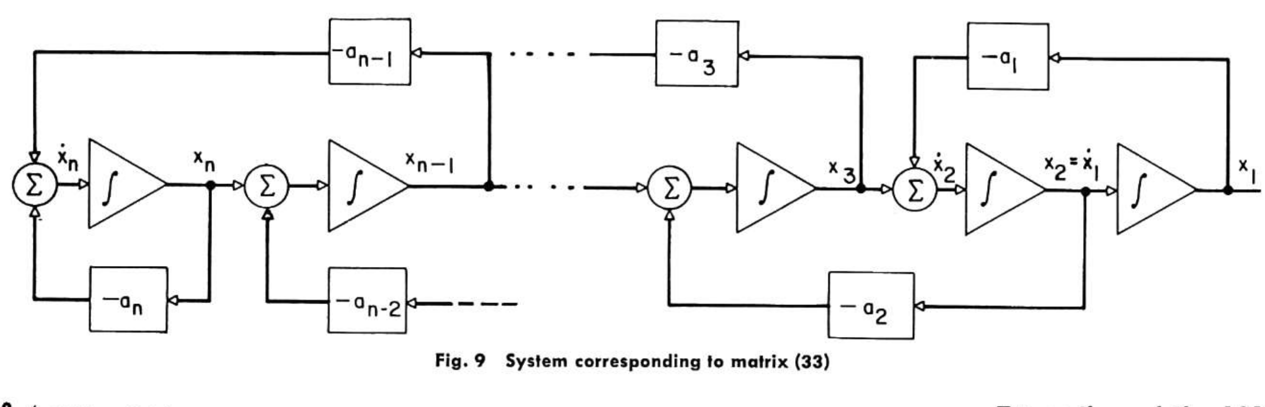


Fig. 9 System corresponding to matrix (33)

$$p_{ii} = a_n a_{n-1} \ldots a_i \qquad (i = 1, \ldots, n)$$

while all other elements are zero. If all the a_i are positive, then $\mathbf{P}$ is clearly positive definite. Let $V(\mathbf{x}) = \|\mathbf{x}\|^2_{\mathbf{P}}$. This function clearly satisfies hypotheses (i-A) and (iv-A) of Corollary 1.2. Further, a simple computation shows:

$$\dot{V}(\mathbf{x}) = \dot{\mathbf{x}}'(\text{grad } V) = 2\mathbf{x}'\mathbf{C}'\mathbf{P}\mathbf{x} = -2a_n x_n^2 \leqq 0$$

which satisfies hypothesis (ii$_1$-A) of Corollary 1.3. To verify hypothesis (ii$_2$-A), it is necessary to show that $\dot{V}(\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0))$ is not identically zero in t for any initial state $\mathbf{x}_0 \neq \mathbf{0}$. Now if $\phi_n^2(t;\ \mathbf{x}_0, t_0)$ is to be identically zero in t, then by inspection of Fig. 9 it follows that $d\phi_n(t;\ \mathbf{x}_0, t_0)/dt$ must be identically zero in t. This in turn requires that $\phi_{n-1}(t;\ \mathbf{x}_0, t_0)$ be identically zero in t. Proceeding in this fashion, we see that every component of $\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0)$ must vanish identically in t, i.e., $\mathbf{x}_0$ is an equilibrium state of (34). Because

$$\det \mathbf{C} = \begin{Bmatrix} -\ a_1a_3 \ldots a_n & n = \text{odd} \\ +\ a_1a_3 \ldots a_{n-1} & n = \text{even} \end{Bmatrix} \neq 0$$

the system (34) has but one equilibrium state which is $\mathbf{x}_e = \mathbf{0}$. Thus $\dot{V}$ does not vanish identically along any motion starting at $\mathbf{x}_0 \neq \mathbf{0}$.

7.2 Stability of Linear Nonstationary Systems. No procedure similar to Corollary 3.1 is available so far which would permit determination of the stability of general linear nonstationary systems in an algebraic way. There is little hope that this state of affairs will change soon. It is therefore necessary to use ingenuity in finding explicit Lyapunov functions. Without discussing this matter in detail (see References [28], [29]) let us outline briefly an obvious procedure.

Take a matrix $\mathbf{P}(t)$ which is a differentiable function of t subject to

$$0 < c_1\mathbf{I} \leqq \mathbf{P}(t) \leqq c_2\mathbf{I} < \infty$$

for all t. Then the function

$$V(\mathbf{x}, t) = \|\mathbf{x}\|^2_{\mathbf{P}(t)} \tag{35}$$

obviously satisfies hypotheses (i), (iii), (iv) of Theorem 1 (in fact, $\alpha(\|\mathbf{x}\|) = c_1\|\mathbf{x}\|^2$, $\beta(\|\mathbf{x}\|) = c_2\|\mathbf{x}\|^2$). If, in addition,

$$-\mathbf{Q}(t) = \mathbf{F}'(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}(t) + \dot{\mathbf{P}}(t) \leqq -c_3\mathbf{I} < 0 \tag{36}$$

then $$\dot{V}(\mathbf{x}, t) = -\|\mathbf{x}\|^2_{\mathbf{Q}(t)}$$

satisfies hypothesis (ii) of Theorem 1, and V defined by (35) is a Lyapunov function.

The problem, of course, is to choose $\mathbf{P}(t)$ in such a way that $\mathbf{Q}(t)$ is positive definite for the largest possible region of parameters $\mathbf{F}(t)$. At first, it is best to fix $\mathbf{Q}$ and use Corollary 3.1 to get an initial guess for $\mathbf{P}$.

Example 8. Time-Varying Network. Consider the time-varying network shown in Fig. 10. The differential equations governing the dynamics are:

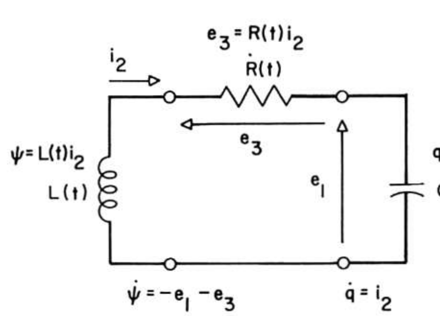


Fig. 10 Time-varying network

$$\left.\begin{aligned} \dot{x}_1 &= L^{-1}(t)x_2 \\ \dot{x}_2 &= -C^{-1}(t)x_1 - R(t)L^{-1}(t)x_2 \end{aligned}\right\} \tag{37}$$

where $x_1 = q$ = charge on capacitor and $x_2 = \psi$ = flux in inductor. Let

$$\mathbf{P}(t) = \begin{bmatrix} R(t) + 2L(t)/R(t)C(t) & 1 \\ 1 & 2/R(t) \end{bmatrix}$$

and it follows immediately from (36) that

$$\mathbf{Q}(t) = \begin{bmatrix} 2C^{-1}(t) & 0 \\ 0 & 2L^{-1}(t) \end{bmatrix} - \dot{P}(t)$$

Assuming

$$\left.\begin{aligned} 0 < \epsilon_1 \leqq L(t) \leqq \nu_1 < \infty \\ 0 < \epsilon_2 \leqq C(t) \leqq \nu_2 < \infty \\ 0 < \epsilon_3 \leqq R(t) \leqq \nu_3 < \infty \end{aligned}\right\} t \geqq t_0$$

the system (37) is asymptotically stable if for all $t \geqq t_0$

$$0 < \epsilon_4 \leqq 1 + \dot{R}(L/R^2 - C/2) + \dot{C}L/RC - \dot{L}/R$$

$$0 < \epsilon_5 \leqq 1 + \dot{R}L/R^2$$

Even in this simple case, it is difficult to get sharp (i.e., almost necessary) sufficient conditions without considering in detail the functions $L(t)$, $C(t)$, and $R(t)$.

7.3 Stability of Nonlinear Systems. In this case, even more than before, intuition must be used to obtain suitable Lyapunov functions. There are a number of well-known examples of how to do this; knowledge of the examples helps to obtain new Lyapunov functions.

Example 9. Linearization About Equilibrium. It is common usage in engineering to consider only small deviations from the operating point (i.e., equilibrium state). This is done by expanding $\mathbf{f}$ in (5-F) in a Taylor series about the operating point and neglecting all higher-order terms.

It is easy to give a rigorous proof of the legitimacy of this procedure by means of the second method. Let $\mathbf{y} = \mathbf{x} - \mathbf{x}_e$ and assume that $\mathbf{f}$ in (5-F) is analytic in a neighborhood of $\mathbf{x}_e$. In that neighborhood $\mathbf{f}$ has a Taylor series which can be written as:

$$\begin{aligned} d\mathbf{y}/dt &= \mathbf{f}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}_e, t) + [\mathbf{F}(t) + \mathbf{G}(\mathbf{y}, t)](\mathbf{x} - \mathbf{x}_e) \\ &= [\mathbf{F}(t) + \mathbf{G}(\mathbf{y}, t)]\mathbf{y} \end{aligned} \tag{38}$$

where $\mathbf{F}(t)$ is the Jacobian matrix of $\mathbf{f}$ evaluated at $\mathbf{x}_e$, t; and $\mathbf{G}(\mathbf{y}, t)$ is a matrix such that $\|\mathbf{G}(\mathbf{y}, t)\mathbf{y}\|/\|\mathbf{y}\|$ tends to 0 with $\|\mathbf{y}\| \to 0$.

Now suppose that

(i) $\|\mathbf{F}(t)\| \leqq c_1$ *for all* t

(ii) *The linearized system*

$$d\mathbf{y}/dt = \mathbf{F}(t)\mathbf{y} \tag{39}$$

is uniformly asymptotically stable.

Then by part (E) of Theorem 3 there is a Lyapunov function $V(\mathbf{y}, t)$ for (39). The derivative of this function along motions of (38) is

$$\dot{V}(\mathbf{y}, t) = -\|\mathbf{y}\|^2_{\mathbf{Q}(t)} + 2\mathbf{y}'\mathbf{P}(t)\mathbf{G}(\mathbf{y}, t)\mathbf{y}$$

By Schwarz's inequality and (16),

$$\leqq -c_9\|\mathbf{y}\|^2 + 2\|\mathbf{P}(t)\mathbf{y}\| \cdot \|\mathbf{G}(\mathbf{y}, t)\mathbf{y}\|$$

By (23), (76), and (80), see Appendix,

$$\leqq -c_9\|\mathbf{y}\|^2 + \frac{c_7^2 c_{10}}{c_8}\|\mathbf{y}\| \cdot \|\mathbf{G}(\mathbf{y}, t)\mathbf{y}\| \tag{40}$$

If we assume further that

(iii) $\|\mathbf{G}(\mathbf{y}, t)\mathbf{y}\|/\|\mathbf{y}\| \to 0$ *uniformly in t with* $\|\mathbf{y}\| \to 0$

then (see Theorem 1) $V(\mathbf{y}, t)$ will be a Lyapunov function also for system (38) in a sufficiently small neighborhood of the origin. Therefore we conclude that:

If the linearized system (39) near equilibrium is uniformly asymptotically stable and is a uniformly good approximation to the original system near $\mathbf{x}_e$, *then* $\mathbf{x}_e$ *of (38) is (locally) uniformly asymptotically stable.*

This result includes in it a method of considering small nonlinear terms in an otherwise linear system. This will be discussed in more detail in Example 15.

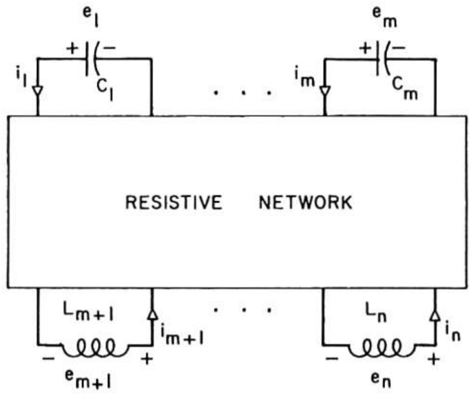


Fig. 11 General RLC network

Example 10. Passive Nonlinear Network. Consider a network of nonlinear inductors, capacitors, and resistors, without internal sources, mutual inductances, or ideal transformers. Let the first m elements be inductors, the next n-m elements capacitors, the rest resistors. See Fig. 11. Let e_j, i_j ($j = 1, \ldots, n$) be the voltages across and the currents through the energy storage elements with polarities as shown in Fig. 11. Using ψ to denote flux and q to denote charge, the mathematical description of the nonlinear inductors and capacitors is

$$i_j = f_j(\psi_j), \qquad e_j = d\psi_j/dt \qquad (j = 1, \ldots, m)$$

$$e_j = f_j(q_j), \qquad i_j = dq_j/dt \qquad (j = m + 1, \ldots, n)$$

Evidently $\psi_1, \ldots, \psi_m, q_{m+1}, \ldots, q_n$ is a natural choice for the state variables $x_1, \ldots, x_n$. The stored energy in the system is given by

$$\begin{aligned} E = V(\mathbf{x}) &= \sum_{j=1}^{n} \int_0^\infty e_j i_j dt \\ &= \sum_{j=1}^{m} \int_0^{\psi_j} f_j(\psi_j)d\psi_j + \sum_{j=m+1}^{n} \int_0^{q_j} f_j(q_j)dq_j \\ &= \sum_{j=1}^{n} \int_0^{x_j} f_j(x_j)dx_j \end{aligned} \qquad (41)$$

Assuming the network has an equilibrium state at $\mathbf{x} = \mathbf{0}$, i.e.

$$f_j(0) = 0 \qquad (j = 1, \ldots, n)$$

(41) will be a Lyapunov function if

$$x_j f_j(x_j) > 0, \qquad \int_0^\infty f_j(x_j)dx_j = \infty \qquad (j = 1, \ldots, n)$$

and
$$\dot{V}(x) = \sum_{j=1}^{n} e_j i_j < 0$$

The latter condition is always satisfied by passive resistive networks; in other words, by networks containing only resistors whose conductance function is $i = g(e)$ such that (i) $g(0) = 0$, (ii) $g(-e) = -g(e)$, (iii) $g(e)$ is monotonically increasing with e. See equations (6) and (6+) of Reference [30]. Thus *every passive network is asymptotically stable in the large.* While this fact may seem to be obvious physically, the simplest rigorous proof is supplied by the second method of Lyapunov.

An interesting and at present largely unsolved problem which arises in connection with nonlinear control systems is the following: Suppose first that all nonlinearities are functions of a single variable only. Let each nonlinearity $g(\xi)$ be replaced formally by ("linearized") $g(\xi) = K(\xi)\xi$ where either

$$K(\xi) = g(\xi)/\xi \qquad (42)$$

or

$$K(\xi) = dg(\xi)/d\xi = g'(\xi) \qquad (43)$$

where $K(\xi)$ is treated formally as a constant. Assuming stationarity, this leads in general to a differential equation of the form

$$d\mathbf{x}/dt = \mathbf{F}(\mathbf{x})\mathbf{x} \qquad (44)$$

Now suppose the matrix $\mathbf{F}(\mathbf{x})$ is asymptotically stable for all $\mathbf{x}$. Can we conclude from this that the original equation $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is globally asymptotically stable? This has been widely conjectured in engineering circles for some time (in case of linearization (42) by Aizerman [31], in case of linearization (43) by Kalman [32]), and is probably true in most cases. However, no rigorous proof is available at present. More exactly, it is not known in precisely what cases and for precisely what reasons the conjectures fail. There are, however, a number of interesting partial results.

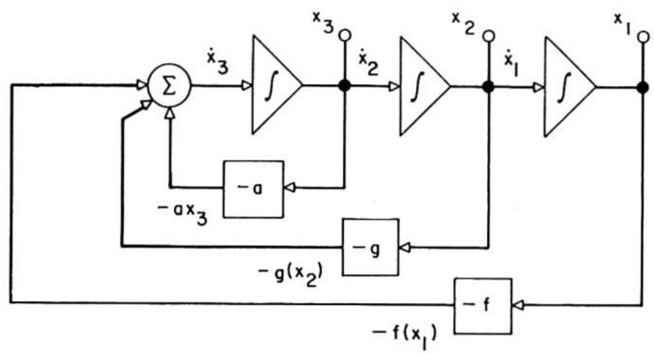


Fig. 12 System corresponding to Barbashin's theorem

Example 11. Stability of a Third-Order Nonlinear System. Barbashin [33, 34] considers the following system (see Fig. 12):

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= -f(x_1) - g(x_2) - a_2 x_3 \end{aligned}\right\} \qquad (45)$$

where $f(0) = g(0) = 0$ and f, g are differentiable functions.

Using the "second method" in the form of Corollaries 1.2–1.3, Barbashin shows:

The equilibrium state $\mathbf{x}_e = \mathbf{0}$ *of system (45) is asymptotically stable in the large if:*

(i) $a_2 > 0$
(ii) $f(x_1)/x_1 \geqq \epsilon_1 > 0, \quad x_1 \neq 0$
(iii) $a_2 g(x_2)/x_2 - f'(x_1) \geqq \epsilon_2 > 0, \quad x_2 \neq 0$

If we replace $f(x_1)/x_1$ and $f'(x_1)$ by a_0 and $g(x_2)/x_2$ by a_1, then the foregoing conditions become:

(i′) $a_2 > 0$ (ii′) $a_0 > 0$ (iii′) $a_2 a_1 - a_0 > 0$

which are just the Routh-Hurwitz conditions for the linear system

$$\dddot{x}_1 + a_2\ddot{x}_1 + a_1\dot{x}_1 + a_0 x_1 = 0$$

Notice, however, that conditions (i–iii) above involve *both* kinds of linearizations mentioned in the foregoing.

Let

$$F(x_1) = \int_0^{x_1} f(x_1)dx_1 \quad \text{and} \quad G(x_2) = \int_0^{x_2} g(x_2)dx_2$$

and define as a candidate for a Lyapunov function of (45):

$$V(\mathbf{x}) = a_2F(x_1) + f(x_1)x_2 + G(x_2) + (a_2x_2 + x_3)^2/2 \quad (46)$$

In view of (ii), $f'(0) > 0$, and $F(x_1) > 0$ when $x_1 \neq 0$. Further, by (i) and (iii), it follows that $G(x_2) > 0$ when $x_2 \neq 0$. When $x_2 = 0$ (46) is obviously nonnegative and 0 only if $x_1 = x_3 = 0$. When $x_2 \neq 0$, the first three terms in (46) can be written as

$$\frac{[2G(x_2) + f(x_1)x_2]^2 + 4\int_0^{x_1} f(x_1) \int_0^{x_2} [a_2g(x_2)/x_2 - f'(x_1)]x_2dx_2dx_1}{4G(x_2)}$$

The first term in the foregoing is always positive. By (iii) the inner integral above is positive, and therefore by (ii) the outer integral is nonnegative. Putting such facts together shows that V is positive unless $\mathbf{x} = \mathbf{0}$. Because all integrands in question are uniformly bounded from below by (ii) and (iii), it follows further that $V(\mathbf{x}) \to \infty$ when $\|\mathbf{x}\| \to \infty$. Hence V satisfies conditions (i-A), (iii-A) and (iv-A) of Corollary 1.2. After a short calculation, it is found that

$$\dot{V}(\mathbf{x}) = -[a_2g(x_2)/x_2 - f'(x_1)]x_2^2$$

By (iii), $\dot{V}(\mathbf{x}) = 0$ only if $x_2 = 0$. Using exactly the same argument as in Example 7, it follows that $\dot{V}(\boldsymbol{\phi}(t;\ \mathbf{x}_0, 0))$ does not vanish identically along any trajectory starting at $\mathbf{x}_0 \neq \mathbf{0}$. This verifies hypotheses (ii$_1$-A), (ii$_2$-A) of Corollary 1.3 which completes the proof that (46) is a Lyapunov function for (45).

Perhaps the best general result on stability of nonlinear systems (not restricted by dimensionality as in the foregoing example) is the following:

Theorem 4. (Krasovskii [35]) *Consider the continuous-time, free, stationary dynamic system:*

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}) \qquad (\mathbf{f}(\mathbf{0}) = \mathbf{0}) \qquad \text{(5-FS)}$$

ASSUME *that* $\mathbf{f}$ *has continuous first partial derivatives and that its Jacobian matrix* $\mathbf{F}(\mathbf{x}) = [\partial f_i/\partial x_j]$ *satisfies the condition* ($\epsilon > 0$ *is an arbitrary constant*):

(i) $\hat{\mathbf{F}}(\mathbf{x}) = [\mathbf{F}(\mathbf{x}) + \mathbf{F}'(\mathbf{x})] \leqq -\epsilon\mathbf{I} < 0$

THEN *the equilibrium state* $\mathbf{x}_e = \mathbf{0}$ *of the system* (5-FS) *is asymptotically stable in the large and*

$$V(\mathbf{x}) = \|\mathbf{f}(\mathbf{x})\|^2$$

is one of its Lyapunov functions.

PROOF. Computing $\dot{V}$, we find that

$$\begin{aligned} \dot{V}(\mathbf{x}) &= \dot{\mathbf{x}}'(\text{grad } V) \\ &= 2\sum_{j=1}^{n} \dot{x}_j \sum_{i=1}^{n} f_i(x)\,\frac{\partial f_i(x)}{\partial x_j} \\ &= 2\mathbf{f}'(\mathbf{x})\mathbf{F}(\mathbf{x})\dot{\mathbf{x}} \\ &= \|\mathbf{f}(\mathbf{x})\|^2_{\hat{\mathbf{F}}(\mathbf{x})} \leqq -\epsilon\|\mathbf{x}\|^2 < 0 \end{aligned}$$

the last step following by hypothesis (i). Thus $\dot{V}$ is negative definite.

It remains to show that V is positive definite and tends to ∞ with $\|\mathbf{x}\| \to \infty$.

Let $\mathbf{c}$ be any constant, nonzero vector. The set of vectors $\{\alpha\mathbf{c};\ 0 \leqq \alpha \leqq 1\}$ is a straight line (ray) connecting the origin with $\mathbf{c}$. Integrating along this ray, we have the identity

$$f_i(\mathbf{c}) = \sum_{j=1}^{n}\left[\int_0^1 c_j\,\frac{\partial f_i(\alpha\mathbf{c})}{\partial x_j}\,d\alpha\right] \qquad (46)$$

Suppose that $\mathbf{f}(\mathbf{c})$ is $\mathbf{0}$ for some $\mathbf{c} \neq \mathbf{0}$. Then obviously

$$\begin{aligned} 0 = \mathbf{c}'\mathbf{f}(\mathbf{c}) &= \sum_{i=1}^{n} c_i f_i(\mathbf{c}) \\ &= \int_0^1 \left[\sum_{i,j=1}^{n} c_i\,\frac{\partial f_i(\alpha\mathbf{c})}{\partial x_j}\,c_j\right]d\alpha \\ &\leqq -\epsilon\|\mathbf{c}\|^2/2 < 0 \end{aligned}$$

which is a contradiction. Hence V is positive definite. The preceding argument shows also that $\mathbf{x}'\mathbf{f}(\alpha\mathbf{x}) \to -\infty$ with $\alpha \to \infty$, for any fixed vector $\mathbf{x} \neq \mathbf{0}$. But this can happen only if at least one component of $\mathbf{f}(\mathbf{x})$ tends to infinity in absolute value as $\|\mathbf{x}\|$ tends to infinity which completes the proof.

The most obvious illustration of this result is the following:

Example 12. Passive Linear Stationary System. In Example 10, in the linear case, one can define the state variables in such a way that the infinitesimal transition matrix $\mathbf{F}$ is negative definite. This is due to the fact that the terminal impedance matrix of an n-port network containing only linear, passive resistors is negative definite. Then stability is immediate, by Theorem 4. In this special case, therefore, the Routh-Hurwitz inequalities are actually identical with the ones obtained by testing for negative definiteness.

A less trivial example is:

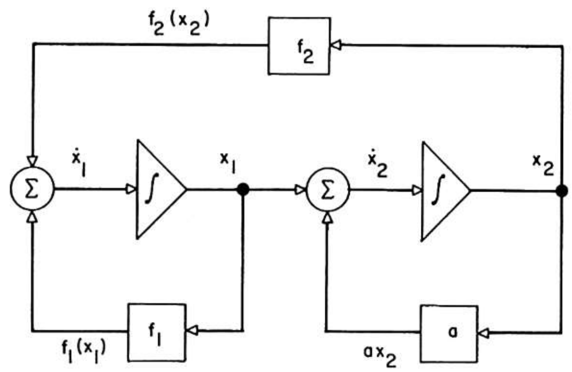


Fig. 13 System illustrating Krasovskii's theorem

Example 13. Second-Order System With Two Nonlinearities Fig. 13 illustrates the system

$$\begin{aligned} \dot{x}_1 &= f_1(x_1) + f_2(x_2) \\ \dot{x}_2 &= x_1 + ax_2 \end{aligned} \qquad (47)$$

where $f_1(0) = f_2(0) = 0$; f_1 and f_2 are differentiable functions.

Here

$$\hat{\mathbf{F}}(\mathbf{x}) = \begin{bmatrix} 2f_1'(x_1) & 1 + f_2'(x_2) \\ 1 + f_2'(x_2) & 2a \end{bmatrix}$$

To satisfy the hypothesis of Theorem 4 we must have

$$4af_1'(x_1) - [1 + f_2'(x_2)]^2 \geqq \epsilon^2 > 0 \quad \text{for all } \mathbf{x}$$
$$f_1'(x_1) \leqq -\epsilon < 0 \text{ for all } x_1 \tag{48}$$

These requirements are considerably more restrictive than the formal Routh-Hurwitz conditions which would be

$$af_1(x_1)/x_1 - f_2(x_2)/x_2 \geqq \epsilon^2 > 0 \quad \text{for all } x_1 \neq 0, \quad x_2 \neq 0$$
$$f_1(x_1)/x_1 + a \leqq -\epsilon < 0 \quad \text{for all } x_1 \neq 0 \tag{49}$$

or, using the other type of linearization,

$$af_1'(x_1) - f_2'(x_2) \geqq \epsilon^2 > 0 \quad \text{for all } \mathbf{x}$$
$$f_1'(x_1) + a \leqq -\epsilon < 0 \quad \text{for all } x_1 \tag{50}$$

where (50) evidently implies (49). Using a special Lyapunov function, Krasovskii [36] was able to show that the asymptotic stability in the large of the origin of the system (47) is guaranteed even when conditions (48) are replaced by the *first* of conditions (50) and the *second* of conditions (49). On the other hand, as Krasovskii also shows, conditions (49) are not sufficient even for local asymptotic stability.

It is not a simple matter to obtain Routh-Hurwitz-like conditions for nonlinear systems since it is not clear what *kind* of linearization is the most natural one. Much more mathematical research remains to be done on these problems.

8—Estimation of Transient Behavior

If the stability of a given system has been established by means of a Lyapunov function, the latter can be used to estimate the rapidity of transient response, effect of perturbations or variation in parameters, etc. All of these applications stem from regarding the Lyapunov function as a measure of distance in the state space.

Consider the obvious inequality

$$\dot{V}(\mathbf{x}, t) = [\dot{V}(\mathbf{x}, t)/V(\mathbf{x}, t)]V(\mathbf{x}, t)$$
$$\leqq -\eta V(\mathbf{x}, t) \tag{51}$$

where η is the minimum of the ratio $-\dot{V}(\mathbf{x}, t)/V(\mathbf{x}, t)$ in some region of the state space excluding the origin. For instance, in case of uniform asymptotic stability in the large,

$$\eta = \operatorname*{Inf}_{\mathbf{x}} \left\{ \frac{\gamma(\|\mathbf{x}\|)}{\beta(\|\mathbf{x}\|)};\ 0 < \|\mathbf{x}\| \leqq r \right\} \geqq 0\ ; \tag{52}$$

if r is chosen so that $\beta(\|\mathbf{x}_0\|) < \alpha(r)$ then the motion will never leave the region $\|\mathbf{x}\| \leqq r$ so that the inequality (51) is applicable for all $t \geqq t_0$. By elementary calculation, (51) leads to

$$V(\boldsymbol{\phi}(t;\ \mathbf{x}_0, t_0), t) \leqq \exp\,[-\eta(t - t_0)]V(\mathbf{x}_0, t_0) \tag{53}$$

Interpreting V as the distance from the origin, it is clear that (53) gives an estimate of how fast equilibrium is approached.

In fact, η^{-1} can be interpreted as the largest time-constant over a certain region in state space and may be regarded as a figure of merit of the system. A number of points should be observed in this connection:

(*a*) If $\dot{V}$ is only semidefinite, we may be able to conclude asymptotic stability by means of Corollary 1.3. But then $\eta = 0$ and we cannot use (51) to estimate the rate of approach to equilibrium—in fact, then (51) merely implies stability. Fortunately, by Theorem 2, one can always find, in principle, a Lyapunov function with negative definite $\dot{V}$ in case of uniform asymptotic stability. Another situation where semidefinite $\dot{V}$ leads to difficulty occurs in the next section.

(*b*) It may be useful to consider the values of η over several different regions of state space. For instance, in Example 2, $\eta = 2aV$ by equation (4). Since $V \to 0$ with $\mathbf{x} \to \mathbf{0}$, it follows that, if η is defined by (52), it is zero. On the other hand, in any annular region $r_1 \leqq \|\mathbf{x}\| \leqq r_2$, $\eta = 2ar_1^2$.

(*c*) The specific value of η depends on the Lyapunov function chosen. Therefore one may want to choose V so as to maximize η. Little is known at present as to how this can be done.

Example 14. Calculation of η in Linear Stationary Case. Consider a free, linear, stationary, asymptotically stable system. Let $\mathbf{Q}$ be an arbitrary symmetric, positive-definite matrix and find the corresponding symmetric, positive-definite matrix $\mathbf{P}$ by Corollary 3.1. Then

$$V(\mathbf{x}) = \|\mathbf{x}\|^2_{\mathbf{P}}, \qquad \dot{V}(\mathbf{x}) = -\|\mathbf{x}\|^2_{\mathbf{Q}} \tag{54}$$

A definition of η analogous to (52) is then

$$\eta = \operatorname*{Min}_{\mathbf{x}}\{\|\mathbf{x}\|^2_{\mathbf{Q}};\ \|\mathbf{x}\|^2_{\mathbf{P}} = 1\} \tag{55}$$

and now the minimization can be performed with the help of the usual Lagrange multiplier technique. Let Γ be the Lagrange multiplier; then minimizing $\mathbf{x}'(\mathbf{Q} - \Gamma\mathbf{P})\mathbf{x}$ with respect to $\mathbf{x}$ implies that the minimum occurs at a value $\mathbf{x}^*$ of $\mathbf{x}$ such that

$$(\mathbf{Q} - \Gamma\mathbf{P})\mathbf{x}^* = \mathbf{0} \tag{56}$$

Therefore

$$\mathbf{x}^{*\prime}\mathbf{Q}\mathbf{x}^* = \Gamma\mathbf{x}^{*\prime}\mathbf{P}\mathbf{x}^* = \Gamma > 0$$

which is a minimum if Γ is. But, by (56), Γ is an eigenvalue of the matrix $\mathbf{QP}^{-1}$. Hence

$$\eta = \lambda_{\min}(\mathbf{QP}^{-1}) \tag{57}$$

The inputs or perturbations acting on a control system are frequently not known accurately, although some estimate of their maximum amplitude may be available. Similarly, the function $\mathbf{f}$ may not be known exactly, or may change due to parameter variations. Both effects may be represented by adding a term on the right-hand side of (5-F). In such cases, there may be no asymptotic stability but one can ask the question, "What is the smallest spherical neighborhood about the origin which the motions of the perturbed system are *sure* to enter?" The radius of this neighborhood depends on the bounds of the perturbation and the parameter variation and can be easily estimated using the "second method."

Write (5-F) in the form

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, t) + \mathbf{g}(\mathbf{x}, t),$$

where

$$\|\mathbf{g}(\mathbf{x}, t)\| \leqq c_0 < \infty \tag{58}$$

Let $V(\mathbf{x}, t)$ be a Lyapunov function for the system (5-F) in accordance with Theorem 1, with derivative $\dot{V}(\mathbf{x}, t)$. Then the derivative of V for the system (58) is

$$\dot{V}(\mathbf{x}, t) + \mathbf{g}'(\mathbf{x}, t)[\mathbf{grad}\ V(\mathbf{x}, t)]$$
$$\leqq -\gamma(\|\mathbf{x}\|) + c_0\|\mathbf{grad}\ V(\mathbf{x}, t)\| \tag{59}$$

If V and γ are given in manageable form, one can then estimate a constant $c_1(c_0)$ such that if $\|\mathbf{x}\| > c_1$, then expression (59) is negative.

Example 15. Effect of Perturbations and Parameter Variations in the Linear Stationary Case. Consider the system

$$d\mathbf{x}/dt = \mathbf{F}\mathbf{x} + \mathbf{G}(\mathbf{x})\mathbf{x} + \mathbf{u}(t) \tag{60}$$

where

$$\|\mathbf{G}(\mathbf{x})\| \leqq c_0$$
$$\|\mathbf{u}(t)\| \leqq c_1$$

Assume that $\mathbf{F}$ is asymptotically stable and define a Lyapunov function by means of Corollary 3.1, as in Example 14, equation (54). Then

$$\dot{V}(\mathbf{x}, t) = -\|\mathbf{x}\|^2_{\mathbf{Q}} + 2[\mathbf{x}'\mathbf{G}'(\mathbf{x}) + \mathbf{u}'(t)]\mathbf{P}\mathbf{x} \qquad (61)$$

By Schwarz's inequality,

$$|\mathbf{x}'\mathbf{G}'(\mathbf{x})\mathbf{P}\mathbf{x}| \leq c_0\|\mathbf{x}\|\cdot\|\mathbf{P}\mathbf{x}\|$$

and

$$|\mathbf{u}'(t)\mathbf{P}\mathbf{x}| \leq c_1\|\mathbf{P}\mathbf{x}\|$$

Hence (61) becomes

$$\dot{V}(\mathbf{x}, t) \leq -\|\mathbf{x}\|^2_{\mathbf{Q}} + 2c_0\|\mathbf{x}\|\cdot\|\mathbf{P}\mathbf{x}\| + 2c_1\|\mathbf{P}\mathbf{x}\| \qquad (62)$$

Now observe that (proof as in Example 14)

$$\|\mathbf{P}\mathbf{x}\|^2 = \|\mathbf{x}\|^2_{\mathbf{P}^2} \leq \lambda_{\max}(\mathbf{P})\|\mathbf{x}\|^2_{\mathbf{P}}$$

Using this, it follows,

$$\dot{V}(\mathbf{x}, t) \leq [-\eta + 2c_2(c_0)]V + 2c_3(c_1)V^{1/2} \qquad (63)$$

where η is as defined in Example 14,

$$c_2(c_0) = c_0\lambda_{\max}{}^{1/2}(\mathbf{P})/\lambda_{\min}{}^{1/2}(\mathbf{P})$$

and

$$c_3(c_1) = c_1\lambda_{\max}{}^{1/2}(\mathbf{P})$$

Therefore if

$$\|\mathbf{x}\| \geq \|\mathbf{x}\|_{\mathbf{P}}/\lambda_{\max}{}^{1/2}(\mathbf{P})$$
$$\geq 2c_1[\eta - 2c_2(c_0)]^{-1} \qquad (64)$$

then $\dot{V}$ is negative.

The quantity $[\eta - 2c_2(c_0)]^{-1}$ can be interpreted as the upper bound on the time constant taking into account the parameter variations. The upper bound of the radius of the region which the motions are sure to enter is proportional to this quantity and to the bound on the noise $\mathbf{u}(t)$.

9—Relations With System Optimization

An important aspect of the "second method" of Lyapunov is its close relationship with problems in the dynamic optimization of control systems. This interrelation has not been explored so far in detail even in the Russian literature and provides much incentive for future research.

Briefly, the situation is as follows. Let $\rho(\mathbf{x})$ be a continuous, nonnegative function which serves as the error criterion of a regulator whose purpose is to maintain the system at all times as close as possible to the equilibrium state $\mathbf{x}_e = \mathbf{0}$. Thus $\rho(\mathbf{0}) = 0$. For instance, one could take $\rho(\mathbf{x}) = x_1^2$, or $\rho(\mathbf{x}) = 1 - \exp(-x_1^2)$, or $\rho(\mathbf{x}) = \sum_{i=1}^{n} |x_i|$, etc.

Now define the *performance index* of the system as the integrated error criterion (assuming stationarity):

$$V(\mathbf{x}) = \int_0^\infty \rho(\boldsymbol{\phi}(\tau;\ \mathbf{x}, 0))d\tau \qquad (65)$$

This integral is a positive number and can serve as a figure of merit of the system. (When the integral is infinite, it is useless; this can be easily avoided by replacing the upper limit of integration by T, where T denotes an interval of time over which the behavior of the system is of interest.)

If the regulator contains adjustable parameters $\mathbf{p}$ these can be picked in such a way that the value of $V(\mathbf{x}, \mathbf{p})$ is minimized with respect to those parameters for one or more initial state $\mathbf{x}$. This formulation of the optimization problem contains the well-known Wiener-Hall theory [37].

Now the mere fact that the integral (65) is minimized by a physically realizable system does not imply at all that this system is also asymptotically stable. The latter will be the case, however, if $V(\mathbf{x})$ is a suitable Lyapunov function.

To be more precise, consider only linear stationary systems (5-LS). Because ρ is nonnegative and the integrand is continuous in τ, the integral in (65) can be zero only if

$$\rho(\boldsymbol{\phi}(\tau;\ \mathbf{x}, 0)) = 0 \quad \text{for all } \tau \geq 0 \qquad (66)$$

Assume now that $\rho(\mathbf{x})$ is positive definite. Then (66) implies, using also linearity and stationarity:

$$\boldsymbol{\phi}(\tau;\ \mathbf{x}, 0) = (\exp \mathbf{F}\tau)\mathbf{x} = \mathbf{0},\ \tau \geq 0$$

Since the transition matrix is nonsingular, this implies $\mathbf{x} = \mathbf{0}$, which shows that V is positive definite.

Assuming further that V is *finite* in some neighborhood $\|\mathbf{x}\| \leq r$ of the origin, a simple calculation (as in (24)) shows that in the same neighborhood

$$\dot{V}(\mathbf{x}) = -\rho(\mathbf{x}) < 0 \quad \text{when } \mathbf{x} \neq \mathbf{0}$$

i.e., $\dot{V}$ is negative definite. Therefore under the assumptions V satisfies requirements (i-A), (ii-A) of Corollary 1.2 so that (5-LS) is asymptotically stable; by linearity, it is then also asymptotically stable in the large so that the restriction about a finite neighborhood of the origin made above is immaterial. Thus we have obtained the following simple but useful result [16]:

Theorem 5. (Kalman) *Consider a free, linear, stationary dynamic system with an equilibrium state at the origin and* ASSUME

(i) *The error criterion $\rho(\mathbf{x})$ is positive definite and $\rho(\mathbf{0}) = 0$;*

(ii) *The performance index $V(\mathbf{x})$ defined by (65) is finite in some neighborhood of the origin.*

THEN $\mathbf{x}_e = \mathbf{0}$ *is asymptotically stable.*

The assumptions on $\rho(\mathbf{x})$ could be considerably weakened, of course; the extension of the theorem to the linear nonstationary case is completely straightforward.

The importance of this theory is emphasized by the following facts:

(*a*) In the conventional theory of optimization the assumption of positive definiteness of ρ is missing. For this reason, the theory in its classical form is unsatisfactory since there is no proof in general that the optimized system is stable. In the simple cases usually discussed in textbooks, this omission is actually not serious due to the *accidental* fact (cf. here Examples 7 and 11) that often even positive semidefinite error criteria imply that $V(\mathbf{x})$ is a Lyapunov function since the error criterion does not vanish identically along any trajectory (Corollary 1.3). But this is much less likely to happen in systems with several inputs and outputs—one of the reasons why relatively little progress has been made in the theory of multiloop systems. These matters are discussed in detail in a recent paper of Kalman [16, 41].

(*b*) If $V(\mathbf{x})$ can be made finite by *some* choice of the parameters of control system, then it follows that the optimal system will also be stable since $V_{\text{opt}}(\mathbf{x}) \leq V(\mathbf{x})$. But if the open-loop system is unstable, it is by no means obvious that $V_{\text{opt}}(\mathbf{x})$ can be made finite for all $\mathbf{x}$ and thereby stabilize the system. In fact, in some nonlinear systems [38–40], one can at best get local stability in this case. However, Kalman has shown [16, 41] that, if the definition of the control problem is physically meaningful at all, then in the *linear* case an originally unstable system can always be stabilized by a suitably designed controller.

(*c*) Note that Theorem 5 is not based on any special assumptions concerning the error criterion ρ, for instance, that it should be a quadratic function of the state. From the point of view of getting general results, the particular form of ρ is unimportant; unfortunately, however, the problem of minimizing $V(\mathbf{x}, \mathbf{p})$ is usually difficult to carry out and leads to a nonlinear feedback system whenever ρ is not quadratic.

(*d*) By minimizing $V(\mathbf{x}, \mathbf{p})/\rho(\mathbf{x})$ with respect to $\mathbf{p}$, one automatically obtains the best possible upper bound on the time constant of the system as discussed in Section 8.

(*e*) When $\rho(\mathbf{x}) = x_1^2$, the classical method of evaluation of (65) in the linear stationary case is based on the identity

$$\int_0^\infty x_1^2(\tau)d\tau = \frac{1}{2\pi}\int_{-i\infty}^{i\infty} X_1(i\omega)X_1(-i\omega)d\omega$$

where $X_1(i\omega)$ is the Fourier transform of $x_1(t)$. The right-hand side of the integral is then evaluated by means of the calculus of residues; the resulting formulas have been tabulated [37]. Some may regard this as a clever application of the calculus of residues. However, the same task can be solved in a much more general and elementary way by applying Corollary 3.1, i.e., by formulating the problem in the "right" way from the stated point of view and then solving the linear algebraic equations (19). Thus Corollary 3.1 can be used not only as a method of checking stability but also for system optimization. The following elementary example, suggested by Herschel [42], illustrates this:

Example 16. Conventional Optimization of Second-Order Stationary Systems. Consider the second-order linear stationary system in the output x_1:

$$\ddot{x}_1 + 2\zeta\dot{x}_1 + x_1 = 0$$

or

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= -x_1 - 2\zeta x_2\end{aligned} \tag{67}$$

where ζ is an arbitrary constant.

The problem is to minimize the performance index

$$V(\mathbf{x}, \zeta) = \|\mathbf{x}\|^2_{\mathbf{P}} = \int_{t_0}^\infty \|\boldsymbol{\phi}(\tau;\ \mathbf{x}, t_0)\|^2_{\mathbf{Q}}d\tau$$

where

$$\mathbf{Q} = \begin{bmatrix}1 & 0\\ 0 & \mu\end{bmatrix};\quad \mu > 0$$

Using Corollary 3.1, the solution of equation (19) is

$$\mathbf{P}(\zeta) = \begin{bmatrix}\zeta + \dfrac{1+\mu}{4\zeta} & {}^1/_2\\ {}^1/_2 & \dfrac{1+\mu}{4\zeta}\end{bmatrix}$$

or

$$V(\mathbf{x}, \zeta) = \zeta x_1^2 + \left[\frac{1+\mu}{4\zeta}\right](x_1^2 + x_2^2) + x_1x_2 \tag{68}$$

Suppose now that the constant $\zeta(\mu)$ is chosen in such a way that V is a minimum when $x_2 = 0$. This leads to

$$\zeta^*(\mu) = (\sqrt{1+\mu})/2$$

as the optimal choice of ζ.

Notice that $\zeta^*(0) = 0.5$, a well-known result. The objection has been frequently raised that this type of optimization, (i.e., using $\rho(\mathbf{x}) = x_1^2$) yields a system which is too oscillatory. The usual "rule of thumb" is given to be $\zeta \cong 0.7$. This value is given by $\zeta^*(1) = 1/\sqrt{2} = 0.707$. Thus the use of a *positive-definite* error criterion answers some of the objections raised against optimization based on the integrated squared error.

If ζ could be varied instantaneously, then we could just as well write system (67) in the form

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= -x_1 + u_2(t)\end{aligned}$$

where $u_2(t) = \zeta(t)x_2$. If there are no limits on $\zeta(t)$ then there are also no limits on $u_2(t)$ and then Kalman's theory of optimization [16] (see also reference [15]) applies. It follows that $V(\mathbf{x})$ for this system can be made arbitrarily small by letting $\mathbf{u}(t)$ be some linear function of the state:

$$u_2^*(t) = a_1x_1(t) + a_2x_2(t)$$

and therefore we would have also

$$\zeta^*(t) = a_1x_1(t)/x_2(t) + a_2$$

But even if ζ must be a constant always, it follows that the foregoing optimization procedure is not satisfactory, because one should conceivably still select ζ *as a function of the initial state* $\mathbf{x}$. For instance, if the initial state is $x_1 = 0$ and x_2 arbitrarily, then the optimum value of ζ for this initial state is ∞.

Example 17. Strict Optimization of Second-Order Stationary Systems. A more satisfactory approach to the problem of the preceding example is to regard the optimization as a (two-person, zero-sum) *game* between Nature and Man. Nature chooses the initial state $\mathbf{x}$; Man chooses a (constant) value of ζ. Nature attempts to maximize and Man to minimize the time constant:

$$\tau = \eta^{-1} = V(\mathbf{x}, \zeta)/\rho(\mathbf{x})$$

It is intuitively obvious and easily proved [43] that if Nature plays first, the *payoff* (i.e., the value of τ) will be at least as great as it would be if Man were to play first. This is expressed by the well-known inequality [43]:

$$\tau_* = \underset{\mathbf{x}}{\text{Max}}\{\underset{\zeta}{\text{Min}}[V(\mathbf{x}, \zeta)/\rho(\mathbf{x})]\} \leqq$$
$$\leqq \underset{\zeta}{\text{Min}}\{\underset{\mathbf{x}}{\text{Max}}[V(\mathbf{x}, \zeta)/\rho(\mathbf{x})]\} = \tau^* \tag{69}$$

Here τ_* is the minimum that Nature is sure to "get" and τ^* is the maximum that Man may have to "pay."

The payoff function (68–69) of this continuous game is *convex* in the *strategies* ζ; for this class of games a fairly complete theory is available (Reference [43], Chap. 12). It can be easily computed that here actually $\tau^* = \tau^*$. This means that the game has a *saddle point*, i.e., a *solution* of the game consists of *pure strategies* $\mathbf{x}^*$ and ζ^* which are the values of $\mathbf{x}$, ζ at which the Min-Max (or Max-Min) of $V(\mathbf{x}, \zeta)/\rho(\mathbf{x})$ is attained. An elementary but somewhat lengthy calculation shows that, when $\mu = 1$,

$$x_1^* = \sqrt{(1+\sqrt{5})}\|\mathbf{x}\|/2 = 0.899\|\mathbf{x}\|$$

$$x_2^* = \sqrt{(3-\sqrt{5})}\|\mathbf{x}\|/2 = 0.437\|\mathbf{x}\|$$

$$\zeta^* = \sqrt{\frac{2}{1+\sqrt{5}}} = 0.786$$

$$\tau_* = \tau^* = {}^1/_2(\zeta^* + 1/\zeta^* + \sqrt{\zeta^{*2}+1}) = 1.665$$

Since the value of the game does not change if $\mathbf{x}$ is replaced by $\alpha\mathbf{x}$ ($\alpha \neq 0$), the optimal strategy $\mathbf{x}^*$ is determined only up to a constant multiplier.

Thus we have a solution of the optimization problem which does not depend on any arbitrary assumptions as to the initial state; τ^* is the minimum of the maximum possible values of the time constant. This agrees well with the "rule of thumb" quoted above.

Such problems have not been considered in the literature so far and it seems to be an open question as to whether the optimal solution of games of this type always results in pure strategies.

10—Applications to Design

The "second method" can be used in a variety of ways in connection with system design. A particular advantage of this way of approaching design is that one has then a direct way of assuring stability. A full discussion of these matters is given in the books of Lur'e [7] and Letov [8] cited before.

To give an idea of the possibilities present, we discuss briefly the problem of linear compensation of control systems subject to saturation.

Consider a plant governed by the equations

$$d\mathbf{x}/dt = \mathbf{F}\mathbf{x} + \mathbf{D}\mathbf{u}(t) \tag{70}$$

where the control variables are subject to the constraints,

$$|u_i(t)| < a_i < \infty \qquad (i = 1, \ldots, m) \tag{71}$$

This is essentially the relay or saturating servo problem. The problem is to return every initial state to the origin.

It is well known [38–40] that if $\mathbf{F}$ has eigenvalues with positive real parts then there are some states which cannot be returned to the origin by *any* control subject to the constraints (71). Hence it is assumed that the free system in (70) is asymptotically stable; i.e., that $Re\ \lambda_i(\mathbf{F}) < 0$ for all i. The case where $\mathbf{F}$ has eigenvalues with zero real parts can be treated similarly, albeit much less simply.

Under these assumptions, consider the problem of optimizing the transient behavior of (70) by suitable choice of $\mathbf{u}(t)$. A rigorous optimization, such as minimization of an integral performance index (65) or taking any initial state to the origin in minimum time, will lead to $\mathbf{u}(t)$ being a *nonlinear* function of the instantaneous state. This problem is not yet completely solved.

The following method, suggested by Bass [44], leads to a very simple design procedure of practical significance.

Choose $\mathbf{Q} > \mathbf{0}$ arbitrarily, and employ Corollary 3.1 to get $\mathbf{P} > \mathbf{0}$. Then $\|\mathbf{x}\|^2_{\mathbf{P}} = V(\mathbf{x})$ is a Lyapunov function for system (70), with $\mathbf{u}(t)$ identically zero. Now choose $\mathbf{u}(t)$ so as to make $\dot{V}$ as negative as possible. From (70)

$$\dot{V}(\mathbf{x}(t), t) = -\|\mathbf{x}(t)\|^2_{\mathbf{Q}} + 2\mathbf{u}'(t)\mathbf{D}'\mathbf{P}\mathbf{x}(t)$$

The second term in this expression will be a minimum whenever each control variable $u_i(t)$ has its maximum magnitude, with a sign opposite to that of the corresponding component of the vector $\mathbf{D}'\mathbf{P}\mathbf{x}(t)$. Hence

$$u_i^*(t) = -a_i \operatorname{sgn} (\mathbf{D}'\mathbf{P}\mathbf{x}(t))_i \qquad (i = 1, \ldots, m) \tag{72}$$

where *sgn* is as defined in (21).

This method of design is not a true optimum since

$$\min_{\mathbf{u}(t)} \{V(\mathbf{x}, \mathbf{u}(t))/[-\dot{V}(\mathbf{x}, \mathbf{u}(t))]\} < V(\mathbf{x}, \mathbf{u}(t))\cdot\{\min_{\mathbf{u}(t)}[1/-\dot{V}(\mathbf{x}, \mathbf{u}(t))]\}$$

On the other hand, the design has the advantage that it leads to a system in which there is only linear feedback preceding the nonlinearity.

The result (72) is deceptively simple and it is important to bear in mind two major mathematical difficulties.

First, if the optimal control defined by (72) is substituted into (70) there is no guarantee that solutions of the differential equation will *exist,* since *sgn* is a discontinuous function. This is a well-known phenomenon in the theory of relay servos which was first discovered by Flügge-Lotz [45] and later studied in considerable detail by André and Seibert [46]. However, the "second method" still applies, after a small modification in the definition of $\dot{V}$; the conclusions so obtained agree with a rigorous treatment of the relay servo problem. The difficulties just mentioned may be avoided by replacing (72) with

$$u_i^* = -a_i \operatorname{sat} k(\mathbf{D}'\mathbf{P}\mathbf{x})_i = u_i^*(k\mathbf{D}'\mathbf{P}\mathbf{x}) \quad (i = 1, \ldots, m) \tag{72-A}$$

where

$$\operatorname{sat} x = \begin{cases} +1, & x > 1 \\ x, & |x| \leqq 1 \\ -1, & x < -1 \end{cases}$$

and letting the positive constant k be as large as desired. This provides an arbitrarily good approximation to *sgn x*.

Second, the design represented by (72-A) is realizable only if all state variables can be *measured.* If this is not the case, the unmeasurable state variables must be "synthesized" from the measurable ones by simulating [12] a part of the dynamic equation (70). The necessary simulation equipment (in essence, a small computer) then becomes part of the controller itself (see Example 18 and Fig. 14). However, it is then not at all obvious as to whether the over-all system, including the simulator, will be asymptotically stable. Moreover, there may be difficulties due to inaccurate knowledge of $\mathbf{F}$ and $\mathbf{D}$.

To examine these questions more fully, let $\mathbf{y}$ denote the measurable and $\mathbf{z}$ the unmeasurable state variables. Then (70) can be written in the form:

$$\begin{bmatrix} d\mathbf{y}/dt \\ d\mathbf{z}/dt \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} + \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix} \mathbf{u}(t)$$

where $\mathbf{F}_{ij}$ and $\mathbf{D}_i$ $(i, j = 1, 2)$ denote submatrices of $\mathbf{F}$ and $\mathbf{D}$ defined in the obvious way. Further, let $\hat{\mathbf{z}}$ denote the "synthetic" state variables, and $\hat{\mathbf{F}}_{ij}$, $\hat{\mathbf{D}}_i$ the best available estimate of the quantities $\mathbf{F}_{ij}$, $\mathbf{D}_i$. Use the tilde to denote the error in these estimates such as $\tilde{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{z}}$, etc. Then the equations of the "simulator" are:

$$d\hat{\mathbf{z}}/dt = \hat{\mathbf{F}}_{21}\mathbf{y} + \hat{\mathbf{F}}_{22}\hat{\mathbf{z}} + \hat{\mathbf{D}}_2\mathbf{u}(t)$$

which contains only known quantities. (Assuming there are no random disturbances, $\mathbf{u}(t)$ is always measurable.)

The form of the optimal control is still represented by (72-A) but now $\mathbf{D}$ is replaced by $\hat{\mathbf{D}}$ and $\mathbf{P}$ by $\hat{\mathbf{P}}$, the latter being calculated via (19) from $\mathbf{Q}$ and $\mathbf{F}$. Then (with some obvious notation conventions) the over-all system equations become:

$$\left.\begin{aligned} &(a)\ d\mathbf{x}/dt = \hat{\mathbf{F}}\mathbf{x} + \hat{\mathbf{D}}\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}(\mathbf{x} + \tilde{\mathbf{z}})) \\ &\qquad\qquad + \tilde{\mathbf{F}}\mathbf{x} + \tilde{\mathbf{D}}\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}(\mathbf{x} + \tilde{\mathbf{z}})) \\ &(b)\ d\tilde{\mathbf{z}}/dt = \hat{\mathbf{F}}_{22}\tilde{\mathbf{z}} + \tilde{\mathbf{F}}_2 \cdot \mathbf{x} \end{aligned}\right\} \tag{73}$$

Equations (73) define a dynamic system in the state space consisting of all ordered pairs of vectors $(\mathbf{x}, \mathbf{z})$. The two equations are loosely coupled due to errors in the prediction of $\mathbf{z}$ and in the knowledge of $\mathbf{F}$ and $\mathbf{D}$. We assert:

Theorem 6. (Kalman) *The free, stationary dynamic system (73) is asymptotically stable in the large, for arbitrary $k \geqq 0$, if the following conditions hold:*

(i) $Re\ \lambda_i(\hat{\mathbf{F}}_{22}) < 0$ *for all i.*
(ii) $\tilde{\mathbf{F}}$, $\tilde{\mathbf{D}}$ *are sufficiently small.*

While the theorem is just what one would expect intuitively, the proof is not trivial because of the nonlinear function $\mathbf{u}^*$. (If $\mathbf{u}^*$ is linear, the proof is an immediate consequence of Theorem 3 and the methods of Example 15).

PROOF. (*a*) Consider first the system (73-*a*), with $\tilde{\mathbf{z}} = \mathbf{0}$. By virtue of (72-A), the following function is clearly positive definite for all $c_0 \geqq 0$, $k \geqq 0$:

$$V(\mathbf{x}) = \mathbf{x}'\hat{\mathbf{P}}\mathbf{x} - c_0\mathbf{x}'\hat{\mathbf{P}}\hat{\mathbf{D}}\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\mathbf{x})$$

By simple but somewhat tedious considerations it is verified that

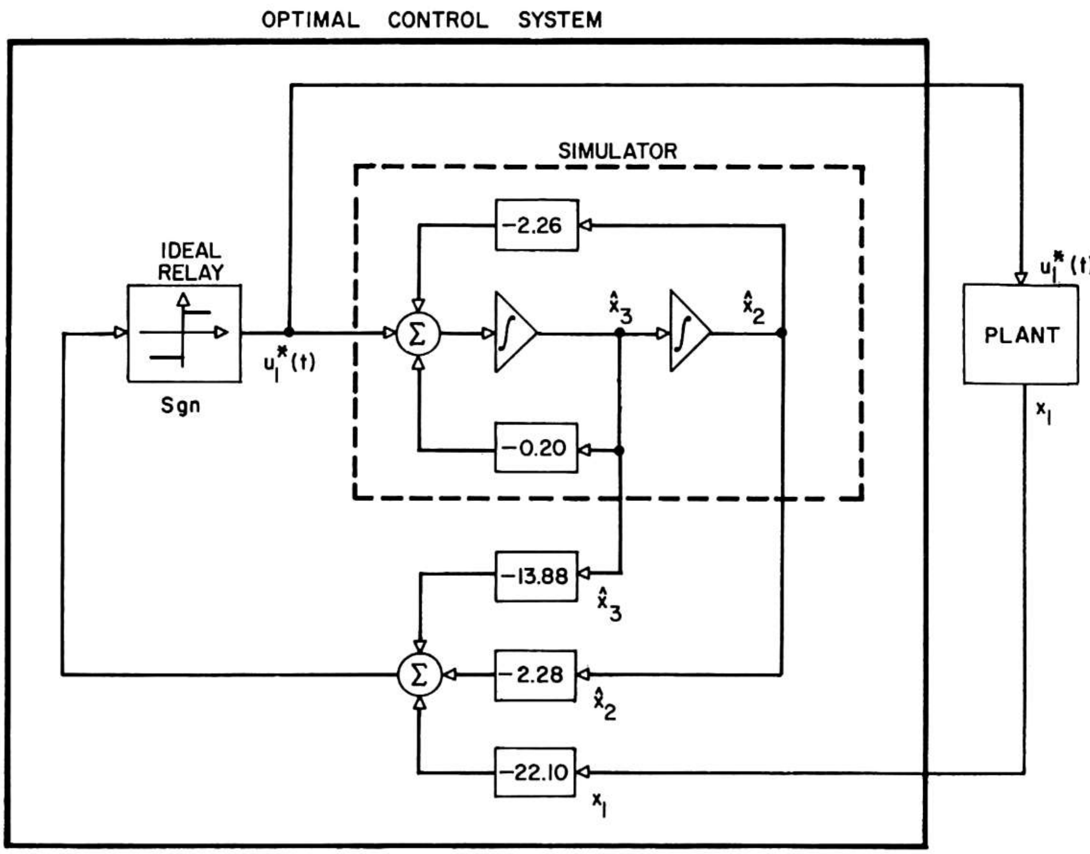


Fig. 14 Design of relay control system

$\dot{V}(\mathbf{x})$ is negative definite, provided that (ii) holds and c_0 is chosen suitably small.

The main steps are as follows. First, note that

$$\dot{V}(\mathbf{x}) \leqq -c_1\|\mathbf{x}\|^2 - c_0c_2\|\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\mathbf{x})\|^2 + (c_0c_3 + c_4)\|\mathbf{x}\|.\|\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\mathbf{x})\|$$

where

$$c_1 = \operatorname*{Min}_{\mathbf{x}} \{\mathbf{x}'(\mathbf{Q} - 2\tilde{\mathbf{F}}'\hat{\mathbf{P}})\mathbf{x};\ \|\mathbf{x}\| = 1\}$$

$$c_2 = \operatorname*{Min}_{\mathbf{x}} \{\mathbf{x}'[(\hat{\mathbf{D}} + \tilde{\mathbf{D}})'\hat{\mathbf{P}}\hat{\mathbf{D}}]\mathbf{x};\ \|\mathbf{x}\| = 1\}$$

$$c_3 = 2\|\mathbf{F}'\hat{\mathbf{P}}\hat{\mathbf{D}}\|$$

$$c_4 = 2\|\hat{\mathbf{P}}\tilde{\mathbf{D}}\|$$

Evidently c_1, c_2 can be made positive and c_4 can be made arbitrarily small by virtue of assumption (ii).

$\dot{V}$ will be negative definite, independently of $k > 0$, if

$$c_0c_1c_2 > (c_0c_3 + c_4)^2/4$$

which can be satisfied by taking c_0, c_4 sufficiently small.

(*b*) Next, consider the system (73), with $\tilde{\mathbf{F}}_{2.} = \mathbf{0}$, in other words, (73-*a*) is coupled to (73-*b*), but not vice versa. Let $\mathbf{M}$ be the matrix which corresponds to $\mathbf{Q} = \mathbf{I}$ and $\hat{\mathbf{F}}_{22}$ through equation (19). By (i), $\mathbf{M}$ will be positive definite.

Now define, as a tentative Lyapunov function for the entire system (73),

$$W(\mathbf{x}, \tilde{\mathbf{z}}) = V(\mathbf{x}) + c_5\|\tilde{\mathbf{z}}\|^2{}_{\mathbf{M}} \tag{74}$$

In view of part (*a*) and noting that $\|\mathbf{u}^*(\mathbf{x} + \tilde{\mathbf{z}})\| < \|\mathbf{u}^*(\mathbf{x})\| + \|\mathbf{u}^*(\tilde{\mathbf{z}})\|$, the derivative of W can be written as

$$\dot{W}(\mathbf{x}, \tilde{\mathbf{z}}) \leqq -c_6\|\mathbf{x}\|^2 - c_7\|\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\mathbf{x})\|^2 - c_5\|\tilde{\mathbf{z}}\|^2 + [c_8\|\mathbf{x}\| + c_9\|\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\mathbf{x})\|)]\|\mathbf{u}^*(k\hat{\mathbf{D}}'\hat{\mathbf{P}}\tilde{\mathbf{z}})\|$$

where

$$0 < c_6 < c_1, \qquad 0 < c_7 < c_0c_2$$

$$c_8 = \|\hat{\mathbf{P}}\mathbf{D}'\|, \qquad c_9 = 2\|\mathbf{D}'\hat{\mathbf{P}}\hat{\mathbf{D}}\|$$

Since evidently

$$\|\mathbf{u}^*(k\mathbf{D}'\mathbf{P}\tilde{\mathbf{z}})\| < 2c_{10}k\|\tilde{\mathbf{z}}\|$$

it follows that $\dot{W}$ is negative definite if the matrix

$$\begin{bmatrix} -c_6 & 0 & c_8c_{10}k \\ 0 & -c_7 & c_9c_{10}k \\ c_8c_{10}k & c_9c_{10}k & -c_5 \end{bmatrix}$$

is negative definite. This will be true if c_5 is large, of the order of k^2. But the choice of $c_5 > 0$ in (74) is arbitrary. Hence for arbitrarily large, but fixed, k there is a Lyapunov function of type (74) such that

$$\dot{W}(\mathbf{x}, \tilde{\mathbf{z}}) < -c_{11}(\|\mathbf{x}\|^2 + \|\tilde{\mathbf{z}}\|^2)$$

(*c*) Finally, consider the case when $\tilde{\mathbf{F}}_{2.} \neq \mathbf{0}$. Since by part (*b*), (73) without *this* perturbation is asymptotically stable, with W and $\dot{W}$ being bounded by *quadratic* forms, hence by the methods of Example 15 it follows that the perturbed system (73) is still asymptotically stable when $\tilde{\mathbf{F}}_{2.}$ is sufficiently small. Q.E.D.

The magnitude of the permissible errors $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{D}}$ can be calculated but this is of little interest except in specific cases.

It must be borne in mind also that asymptotic stability of $\hat{\mathbf{F}}$ by no means implies that of $\hat{\mathbf{F}}_{22}$.

Example 18. Design of a Relay Servo. As a numerical illustration of the foregoing design procedure, consider the plant specified by

$$\mathbf{F} = \begin{bmatrix} -0.01 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2.26 & -0.2 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 \\ 0 \\ 2.5 \end{bmatrix}$$

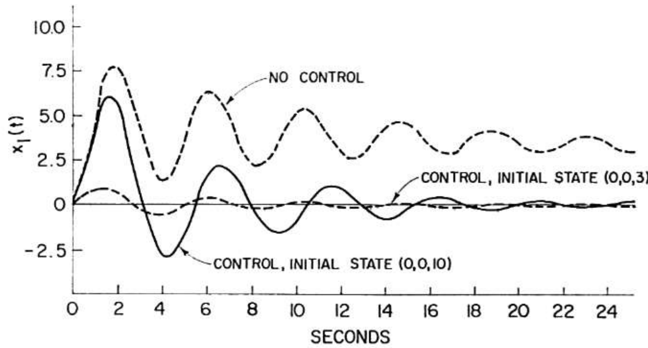


Fig. 15 Response of relay control system

It is easy to see that the transfer function of the plant from $u_1(t)$ to $x_1(t)$ has no zero, a real pole at -0.01 and a pair of complex conjugate poles at $-0.1 \pm i1.5$. Further, assume that $|u_1(t)| \leqq 1$ for all t.

Let $\mathbf{Q} = \mathbf{I}$ and find $\mathbf{P}$ by means of equation (19). Numerical solution leads to

$$\mathbf{P} = \begin{bmatrix} 50.00 & 4.64 & 22.10 \\ 4.64 & 9.71 & 2.28 \\ 22.10 & 2.28 & 13.88 \end{bmatrix}$$

From (72), it follows that the optimal control is given by

$$u_1^*(\mathbf{x}) = \text{sgn}[-22.10x_1 - 2.28x_2 - 13.88x_3]$$

Suppose further that only the output $x_1(t)$ can be measured; the rest of the state variables are to be "synthesized." This leads to the block diagram of Fig. 14. Since $\mathbf{F}_{22}$ is obviously asymptotically stable in this case, it follows that the over-all system shown in the figure is asymptotically stable in the large for small errors in predicting the coefficients of $\mathbf{F}$ and $\mathbf{D}$.

A typical transient response of the uncontrolled plant (i.e., $u_1(t) \equiv 0$ for $t > 0$) is shown in Fig. 15, where $x_1(t)$ is plotted for the initial states $\mathbf{x}_0 = [0, 0, 10]'$, $[0, 0, 3]'$. Note that because of the real pole at -0.01, $x_1(t)$ returns to 0 very slowly. With optimal control and the same initial states $x_1(t)$ goes to zero in about 30 sec even for the largest x_{03}—an improvement by about a factor of 10 over the uncontrolled system.

11—Conclusions

The following aspects of the "second method" of Lyapunov are perhaps most significant in the theory of control systems:

(*a*) The "second method" provides an *abstract* tool for studying stability and transient behavior of dynamic systems without solving the differential equations of these systems.

(*b*) Using a Lyapunov function instead of a norm, one has rigorous methods of analysis of nonlinear systems provided that they are (as in most control-system applications) asymptotically stable in the large. This refutes the oft-repeated misconception in the engineering literature that "nonlinear systems cannot be analyzed by exact methods."

(*c*) The abstract method becomes a concrete one whenever explicit expressions can be found for a Lyapunov function. This is always possible in the linear stationary case, by solving a set of simultaneous algebraic equations of order $n(n + 1)/2$. In the linear nonstationary or nonlinear cases, no straightforward methods are available at present for doing this. Further applied mathematical research should be directed toward developing efficient digital computer programs for finding Lyapunov functions.

(*d*) Given a fixed system, the measure of "goodness" of the Lyapunov function is the number Max $V/(-\dot{V})$ (over some region in the state space); this number can be interpreted as the rigorous upper bound on the time constant of the system.

(*e*) Given a system with free parameters, it is natural to identify $-\dot{V}$ with the error criterion to be used for the system. Then V may be identified with the performance index of the system and is to be minimized by proper choice of the free parameters.

(*f*) The optimized system is stable only if the error criterion does not vanish identically along any trajectory. This added condition is necessary to assure the correctness of the usual Wiener theory of optimization.

(*g*) The "second method" of Lyapunov is more a unifying principle than a method. It does not replace ingenuity. On the other hand, in many cases Lyapunov functions are already available, though unrecognized, in standard results in control theory. These results can then be used in other ways to estimate transient response, effect of random perturbations, and so on.

Acknowledgment

It is a pleasure to acknowledge here our indebtedness to Dr. R. W. Bass of RIAS, whose enthusiasm, persistence, and ever-ready help was a most treasured ally against the intricacies of the second method.

References

1 A. M. Lyapunov, "Problème général de la stabilité du mouvement" (in French), *Ann. Fac. Sci. Toulouse*, vol. 9, 1907, pp. 203–474. Reprinted in Ann. Math. Study No. 17, 1949, Princeton University Press.

2 J. L. Massera, "Contributions to Stability Theory," *Ann. Math.*, vol. 64, 1956, pp. 182–206.

3 W. Hahn, "Theorie und Anwendung der Direkten Methode von Lyapunov," (book; in German), *Erg. der Math.*, No. 22, Springer, 1959, 142 pp.

4 I. G. Malkin, "Theory of Stability of Motion" (book), AEC Translation 3352, Dept. of Commerce, USA, 1958.

5 S. Lefschetz, "Differential Equations: Geometric Theory" (book), Interscience, 1957, 364 pp.; see Chapt. 6, par. 2.

6 S. Lefschetz, "Lyapunov and Stability in Dynamical Systems," *Bol. Soc. Mat. Mex.*, vol. 3, 1958, pp. 25–39.

7 A. I. Lur'e, "Some Nonlinear Problems in the Theory of Automatic Control" (book; Russian), Gostekhizdat, Moscow, 1951; German translation Akademie Verlag 1957; English translation, Her Majesty's Stationery Office, 1957.

8 A. M. Letov, "Stability of Nonlinear Regulating Systems" (book; in Russian), *Gosn. Isdat. Tekhn.-Teor. Lit.*, Moscow, 1955.

9 H. Antosiewicz, "A Survey of Lyapunov's Second Method," in "Contr. to Nonlinear Oscillations" (book), IV, Ann. Math. Study 41, 1958, pp. 147–166, Princeton University Press.

10 V. V. Nemitskii and V. V. Stepanov, "Qualitative Theory of Differential Equations," 2nd ed., Moscow, 1951; English translation of 2nd ed., Princeton University Press, 1960; see chapter 5.

11 E. A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations" (book), McGraw-Hill, 1955; see chapter 3.

12 R. E. Kalman and J. E. Bertram, "General Synthesis Procedure for Computer Control of Single and Multi-Loop Linear Systems," *Trans. AIEE*, vol. 77, II, 1958, pp. 602–609.

13 R. E. Kalman and J. E. Bertram, "A Unified Approach to the Theory of Sampling Systems," *J. Franklin Inst.*, vol. 267, 1959, pp. 405–436.

14 R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," Trans. ASME, vol. 80, 1958, pp. 1820–1826.

15 J. E. Bertram and P. E. Sarachik, "Minimum Control Energy System," being presented at the National Automatic Control Conference, Nov. 1959.

16 R. E. Kalman, "Dynamic Optimization of Linear Control Systems. I. Theory," to appear.

17 L. Markus, "Escape Times for Ordinary Differential Equations," *Rend. Sem. Mat.*, Univ. di Torino, vol. 11, 1951–2, pp. 271–276.

18 J. L. Massera, "On Lyapunov's Conditions of Stability," *Ann. Math.* vol. 50, 1949, pp. 705–721.

19 J. E. Bertram and P. E. Sarachik, "Stability of Circuits With Randomly Time-Varying Parameters," Proc. Int. Symp. on Circuit and Information Theory, Los Angeles, Calif., 1959, pp. 260–270.

20 O. Perron, "The Stability Problem in Differential Equations" (in German), *Math. Zeitsch*, vol. 32 1930, pp. 465–473.
21 L. A. Zadeh, "On the Stability of Linear Time-Varying Systems," *J. Appl. Phys.*, vol. 22, 1952, pp. 402–405.
22 I. G. Malkin, "Certain Questions in the Theory of Stability of Motion in the Sense of Lyapunov" (in Russian), *Sbornik Nauchnykh Trudov*, Kazankogo Aviats. Inst., No. 7, 1937; Am. Math. Soc. Translation No. 20, 1950.
23 H. A. Antosiewicz and P. Davis, "Some Implications of Lyapunov's Conditions of Stability," *J. Rat. Math. and Mech.*, vol. 3, 1954, pp. 447–457.
24 R. E. Bellman, "Kronecker Products and the Second Method of Lyapunov," *Math. Nachrichten*, 1959.
25 R. W. Bass, "Routh-Hurwitz Conditions and the Second Method of Lyapunov," *Proc. Am. Soc.*, to appear.
26 H. R. Schwarz, "A Method for Determining Stability of Matrix Differential Equations" (in German), *Z. angew. Math. Phys.*, vol. 7, 1956, pp. 473–500.
27 A. W. Babister, "Determination of the Optimum Response of Linear Systems," *Q. J. Mech. Appl. Math.*, vol. 10, 1957, pp. 360–368, 504–512; vol. 11 ,1958, pp. 119–128.
28 B. S. Razumikhin, "On the Application of Lyapunov's Methods to Stability Problems," *Prikl. Mat. Mekh.*, vol. 22, 1958, pp. 338–349.
29 Ya. N. Roitenberg, "A Method for the Construction of Lyapunov Functions for Linear Systems With Variable Coefficients," *Prikl. Mat. Mekh.*, vol. 22, 1958, pp. 167–172.
30 G. Birkhoff and J. B. Diaz, "Nonlinear Network Problems," *Q. Appl. Math.*, vol. 13, 1955, pp. 431–443; see also R. J. Duffin, "Nonlinear Networks IIa," *Bull. Am. Math. Soc.*, vol. 53, 1947, pp. 963–971.
31 M. A. Aizerman, "On a Problem Concerning the In-the-Large Stability of Dynamic Systems" (in Russian), *Uspekhi Mat. Nauk*, vol. 4, 1949, pp. 187–188.
32 R. E. Kalman, "On Physical and Mathematical Mechanisms of Instability in Nonlinear Automatic Control Systems," Trans. ASME, vol. 79, 1957, pp. 553–556.
33 E. A. Barbashin, "On the Stability of Solution of a Third-Order Nonlinear Differential Equation," (in Russian), *Prikl. Mat. Mekh.*, vol. 16, 1952, pp. 629–632.
34 M. L. Cartwright, "On the Stability of Solutions of Certain Differential Equations of the Fourth Order," *Q. J. Mech. Appl. Math.*, vol. 9, 1956, pp. 185–194.
35 N. N. Krasovskii, "On the Stability in the Large of a System of Nonlinear Differential Equations" (in Russian), *Prikl. Mat. Mekh.*, vol. 18, 1954, pp. 735–737; also "On Stability Under Large Perturbations" (in Russian), *ibid.*, vol. 21, 1957, pp. 309–319.
36 N. N. Krasovskii, "On the Stability of Solutions of a System of Two Differential Equations" (in Russian), *Prikl. Mat. Mekh.*, vol. 17, 1953, pp. 651–672.
37 G. C. Newton, Jr., L. A. Gould, and J. F. Kaiser, "Analytical Design of Feedback Controls" (book), John Wiley & Sons, New York, N. Y., 1957.
38 R. E. Kalman, "Optimal Control of Saturating Systems by Intermittent Action," IRE WESCON Conv. Record, vol. 1, part 4, 1957, pp. 130–135.
39 C. A. Desoer, "The Bang-Bang Servo Problem Treated by Variational Techniques," *Information and Control*, vol. 2, 1959.
40 J. P. LaSalle, "The Time-Optimal Control Problem," "Contributions to Nonlinear Oscillations" V (book), Annals of Mathematics Studies, Princeton University Press, 1959.
41 R. E. Kalman, "On the General Theory of Control," Moscow Conference on Automatic Control, 1960, to appear.
42 R. Herschel, "On a Generalized Quadratic Optimum," *Regelungstechnik*, vol. 4, 1956, pp. 190–195.
43 J. C. C. McKinsey, "Introduction of the Theory of Games" (book), McGraw-Hill, 1952, 371 pp.
44 R. W. Bass, discussion of a paper by A. M. Letov, Proc. Heidelberg Conf. on Automatic Control ("Regelungstechnik: Moderne Theorien und ihre Verwendbarkeit," by R. Oldenbourg, Munich, 1957), pp. 209–210.
45 I. Flügge-Lotz, "Discontinuous Automatic Control Systems" (book), Princeton University Press, Princeton, N. J., 1953.
46 J. André and P. Seibert, "On Piecewise Linear Differential Equations Occurring in Control Problems" (in German), *Arch. Math.*, vol. 7, 1956, pp. 148–156, 157–164.
47 P. R. Halmos, "Finite-Dimensional Vector Spaces" (book), 2nd Ed., Van Nostrand, 1958.
48 A. S. Householder, "The Approximate Solution of Matrix Problems," *J. Assoc. Comp. Machinery*, vol. 5, 1958, pp. 205–243.
49 R. E. Bellman, "Stability Theory of Differential Equations, McGraw-Hill Book Company, Inc., New York, N. Y., 1953.

APPENDIX

The object here is to present in sufficient detail all the relevant definitions and facts which are needed in the manipulation of norms in applications of the "second method," proofs of theorems, and so on. While these results are very well known, there is no single reference which covers all the necessary points.

The notation at the beginning of Section 4 is used throughout. In calculations, vectors are regarded as matrices with one column. All vectors, matrices, etc. refer to *finite-dimensional* Euclidean spaces.

By a *positive-definite* [*positive semidefinite*] matrix $\mathbf{A}$ we mean any $n \times n$ matrix such that the *quadratic form*

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i,j=1}^{n} x_i a_{ij} x_j$$

is *positive* [*nonnegative*] for all $\mathbf{x} \neq \mathbf{0}$. If $\mathbf{A}$ is positive-definite [semidefinite] then $-\mathbf{A}$ is *negative-definite* [*semidefinite*]. It is not required as a rule that $\mathbf{A}$ be *symmetric* ($\mathbf{A} = \mathbf{A}'$) but note that if $\mathbf{B}$ is any *antisymmetric* matrix $\mathbf{B} = -\mathbf{B}'$ then $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'(\mathbf{A} + \mathbf{B})\mathbf{x}$.

The *eigenvalues* $\lambda_i(\mathbf{A})$ of the $n \times n$ matrix $\mathbf{A}$ are the n-complex roots of the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$. A positive-definite matrix $\mathbf{P}$ always has an inverse, in fact $\lambda_i(\mathbf{P}) > 0$ for all i. Note also that $\lambda_i(\mathbf{AB}) = \lambda_i(\mathbf{BA})$ and $\lambda_i(\mathbf{A}^{-1}) = \lambda_i^{-1}(\mathbf{A})$.

It can be shown that *a matrix* $\mathbf{A}$ *is positive-definite* [*semidefinite*] *if and only if any of the following conditions hold:*

(i) *There is a nonsingular* [*singular*] *matrix* $\mathbf{B}$ *such that* $\mathbf{B}'\mathbf{B} = \mathbf{A}$;

(ii) $\lambda_i(\mathbf{A}) > 0$ [$\lambda_i(\mathbf{A}) \geq 0$] *for all i:*

(iii) *all principal minors of* $\mathbf{A}$ *are positive* [*nonnegative*].

The concept of the norm is similar to that of the absolute value. It is needed in order to be able to make statements about convergence, continuity, etc. in Euclidean space. Standard references are: Halmos [47], chapter 3; Householder [48].

Specifically, a *norm* is a function which assigns to every vector (or point) $\mathbf{x}$ in a given Euclidean space a real number $\|\mathbf{x}\|$ such that

(N_1) $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x}$

(N_2) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}$, $\mathbf{y}$

(N_3) $\|\alpha\mathbf{x}\| = |\alpha|\cdot\|\mathbf{x}\|$ for all $\mathbf{x}$ and complex constant α

(N_4) $\|\mathbf{x}\| = 0$ implies $\mathbf{x} = \mathbf{0}$

If axiom (N_4) is missing, then $\|\mathbf{x}\|$ is a *seminorm*. Examples:

$$(a) \quad \|\mathbf{x}\| = \sum_{i=1}^{n} |x_i|$$

$$(b) \quad \|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = (\mathbf{x}'\mathbf{x})^{1/2}$$

This is the standard definition of the distance of a point in a Euclidean space from the origin, also called *Euclidean norm*.

(c) $\|\mathbf{x}\|_{\mathbf{A}} = (\mathbf{x}'\mathbf{A}\mathbf{x})^{1/2}$ where $\mathbf{A}$ is symmetric and positive definite; called here the *generalized Euclidean norm*.

(d) $\|\mathbf{x}\| = \operatorname{Max}_i \{|x_i|\}$

In explicit calculations, $\|\mathbf{x}\|$ always means the Euclidean norm.

If $\mathbf{A}$ is positive semidefinite then $\|\mathbf{x}\|_{\mathbf{A}}$ is a *seminorm*.

By a *topology* on a mathematical space $\mathcal{X}$ we mean a collection of subsets of $\mathcal{X}$ (called *open sets*) satisfying certain axioms which serve to define abstractly the concept of nearness between elements of $\mathcal{X}$. For a Euclidean space, a topology is usually defined as follows: Call the set of all points $\|\mathbf{x} - \mathbf{y}\| < r$ an *open r-sphere* about $\mathbf{y}$. The topology then consists of all r-spheres ($0 < r < \infty$) about every point $\mathbf{x}$ and their unions and finite intersections. A

Euclidean space together with a given norm is then called a *normed space*.

Convergence and continuity in a space is always defined with respect to a given topology. Two topologies are *equivalent* if every open set of one topology contains an open set and is contained in an open set of the other topology. Two norms in a given finite dimensional Euclidean space define equivalent topologies since it is always possible to find two positive constants c_1, c_2 such that [48]

$$c_1\|\mathbf{x}\|_1 \leqq \|\mathbf{x}\|_2 \leqq c_2\|\mathbf{x}\|_1 \quad \text{for all } \mathbf{x}^{14}$$

For example if $\|\mathbf{x}\|_1$ is defined as in (a) above and $\|\mathbf{x}\|_2$ as in (*b*) above, then it follows from elementary calculations (ref. [11], p. 17)

$$n^{-1/2}\sum_{i=1}^{n}|x_i| \leqq (\mathbf{x}'\mathbf{x})^{1/2} \leqq \sum_{i=1}^{n}|x_i| \tag{75}$$

The method used in Example 14 proves also the following relations between generalized Euclidean norms:

$$\lambda_{\min}^{1/2}(\mathbf{QP}^{-1})\|\mathbf{x}\|_{\mathbf{P}} \leqq \|\mathbf{x}\|_{\mathbf{Q}} \leqq \lambda_{\max}^{1/2}(\mathbf{QP}^{-1})\|\mathbf{x}\|_{\mathbf{P}} \tag{76}$$

($\mathbf{Q}$, $\mathbf{P}$ symmetric and positive-definite).

Closely related to the concept of the norm of a vector is the *norm of a matrix*. This is defined by

$$\|\mathbf{A}\| = \text{Min } K \text{ such that } \|\mathbf{A}\mathbf{x}\| \leqq K\|\mathbf{x}\|$$

For the Euclidean norm, this definition is obviously equivalent to:

$$\|\mathbf{A}\|^2 = \underset{\mathbf{x}}{\text{Max}}\{\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x};\ \mathbf{x}'\mathbf{x} = 1\}$$

[14] Therefore, in questions of convergence, stability, etc., it is immaterial as to what norm is used.

The following inequalities are obvious from the definition

$$\|\mathbf{AB}\| \leqq \|\mathbf{A}\|\cdot\|\mathbf{B}\| \tag{77}$$

$$\|\mathbf{A}'\| = \|\mathbf{A}\| \tag{78}$$

An explicit formula for the Euclidean norm follows at once from the second definition and formula (76):

$$\|\mathbf{A}\|^2 = \lambda_{\max}(\mathbf{A}'\mathbf{A}) \tag{79}$$

Since $\lambda_i(\mathbf{A}^2) = \lambda_i^2(\mathbf{A})$, for a *symmetric* matrix $\mathbf{A}$ the expression for the Euclidean norm is

$$\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A}) = \lambda_{\max}^{1/2}(\mathbf{A}^2) = |\lambda(\mathbf{A})|_{\max} \tag{80}$$

In other cases explicit expressions for the norm are usually not available.

One can easily compute some (more or less pessimistic) upper bounds corresponding to the various norms defined above

$$\|\mathbf{A}\|_a \leqq \sum_{j,i=1}^{n}|a_{ij}|$$

$$\|\mathbf{A}\|_b \leqq \left(\sum_{j,i=1}^{n} a_{ij}^2\right)^{1/2}$$

$$\|\mathbf{A}\|_c \leqq \underset{i}{\text{Max}}\left\{\sum_{j=1}^{n}|a_{ij}|\right\}$$