Automatica, Vol. 12, pp. 519–523. Pergamon Press, 1976. Printed in Great Britain

Brief Paper

Lyapunov Functions for Diagonally Dominant Systems*

J. C. WILLEMS†

Summary—This paper deals with the construction of Lyapunov functions for the finite dimensional linear system $\dot{x} = Ax$ when the entries of the generating matrix A satisfy various conditions requiring dominance of its diagonal elements and nonnegativity of its off-diagonal elements. The particular case in which the system defines a Markov chain is given special attention and it is shown that the results then imply certain inequalities which have an intuitively appealing information theoretic significance.

1. *Introduction*

CONSIDER the autonomous linear system:

$$\Sigma: \dot{x} = Ax$$

with $x = \mathrm{col}(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ (n-dimensional Euclidean space) and $A = (a_{ij})$ $(i, j = 1, 2, \dots, n)$ a real $(n \times n)$ matrix called the *generator* of Σ. This system is said to be *stable* if all its solutions are bounded on $[0, \infty)$ and *asymptotically stable* if all its solutions approach 0 as $t \to \infty$.

One way of verifying the stability of Σ is by constructing a Lyapunov function for it, i.e. by examining the derivative of the real valued function $V(x)$ along Σ. Thus it is well-known that for any matrix C such that (A, C) is observable [1, p. 86] (i.e. such that Rank $[C^T | C^T A^T | \dots | C^T (A^T)^{n-1}] = n$, where T denotes transposition) there exists a positive definite symmetric solution Q to the *Lyapunov equation*

$$A^T Q + QA = -C^T C$$

iff $Re\ \lambda < 0$ for $\lambda \in \sigma(A)$ (Re denotes real part and $\sigma(A)$ denotes the set of eigenvalues of A). Thus if Σ is asymptotically stable this yields a method for constructing quadratic Lyapunov functions $V(x) = \frac{1}{2} x^T Q x$ for Σ on $\mathbb{R}^n$. Other classes of systems for which efficient methods for constructing Lyapunov functions are known are Popov-like systems [2, 3] and the generalization of this methodology to more general "*dissipative*" systems [4].

Of course the construction of Lyapunov functions is not only of interest in order to study the stability of a system. (For the system Σ under consideration there exist indeed much more effective tests than those provided by solving for Q in order to verify stability. For the special classes of systems Σ considered in this paper stability is also essentially obvious). However, the fact that a particular class of functions $V: \mathbb{R}^n \to \mathbb{R}$ are Lyapunov functions may provide a great deal of useful insight in the qualitative behavior of Σ (see, e.g. the results of Section 6), or it may give a candidate for examining the stability of a nonlinear perturbation or analogue of Σ.

*Received 25 August, 1975; revised 9 February, 1976. The original version of this paper was presented at the 6th *IFAC Congress* which was held in Boston/Cambridge Massachusetts during August 1976. The published Proceedings of this IFAC Meeting may be ordered from the ISA–Instrument Society of America, 400 Stanwix Street, Pittxburgh, PA 15222, USA. This paper was recommended for publication in revised form by associate editor B.D.O. Anderson.

†Mathematical Institute, University of Groningen, P.O. Box 800, Groningen, The Netherlands. The final version of this paper was prepared while the author was visiting with the Seminar für Angewandte Mathematik of the ETH, Zürich, Switzerland.

This paper considers the construction of Lyapunov functions for systems Σ in which the entries of the generating matrix A satisfy certain diagonal dominance and/or positivity conditions. System Σ which satisfy such conditions naturally occur in at least two fields: firstly, in economics [5] and secondly in the analysis of Markov chains (see Section 6) where the probabilistic interpretation of the state vector yields the required conditions on A. However, systems as those studied in the paper enter also frequently in engineering applications as well, for example in RC or RL ladder networks, in transistor circuit analysis [6], or in lumped-models of distillation columns [7, 8].

A word about notation. Capital letters usually denote matrices or functions on $\mathbb{R}^n$ whereas lower case letters usually denote vectors or functions on $\mathbb{R}$. For an $(n \times n)$ matrix A, $|A|$ (not to be confused with determinant) denotes the $(n \times n)$ matrix $|A| \triangleq (|a_{ij}|)$ ($|\cdot|$ denotes absolute value). An analogous notation holds for vectors. Further, A_{diag} denotes the diagonal matrix diag (a_{ii}) and $A_{\mathrm{off}} \triangleq A - A_{\mathrm{diag}}$. Finally, $\rho(A)$ denotes the *spectral radius* of the $(n \times n)$ matrix A, i.e. max $|\lambda|$ for $\lambda \in \sigma(A)$. Whenever a condition is assumed to hold for all i, j or σ, say, then the "for all" predicate will for simplicity be deleted.

2. *Dominant, nonnegative, and* M*-matrices*

In this section several special classes of matrices are introduced.

Definition 1. The real $(n \times n)$ matrix $A = (a_{ij})$ is said to be *row dominant* if $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$, *column dominant* if $|a_{jj}| \geq \sum_{i \neq j} |a_{ij}|$, and *doubly dominant* if it is both row and column dominant. If the inequalities are strict then one calls such matrices *strictly* (row, column or doubly) dominant.

Definition 2. A is said to be *nonnegative* if $a_{ij} \geq 0$ and *positive* if $a_{ij} > 0$. A special class of nonnegative matrices are the *stochastic* (resp. *substochastic*) matrices for which $\sum_j b_{ij} = 1$ (resp. ≤ 1) and the *doubly stochastic* (resp. *substochastic*) matrices for which $\sum_j b_{ij} = 1$ (resp. ≤ 1) and $\sum_i b_{ij} = 1$ (resp. ≤ 1). A particular class of doubly stochastic (resp. substochastic) matrices are the *permutation* (resp. *subpermutation*) matrices for which every row and column contains exactly (resp. at most) one element which equals 1, and the remaining elements are 0.

It is well-known [9] that the class of doubly stochastic (resp. substochastic) matrices is the convex hull of the permutation (resp. subpermutation) matrices. This basic fact will be used in Section 3. In Section 4 it will be shown that the matrices introduced in Definition 2 are the matrix exponentials of the following class of matrices:

Definition 3. A is said to be a *generator of nonnegative type* if $a_{ij} \geq 0$ $(i \neq j)$, it is said to have *zero* (resp. *nonpositive*) *column excess* if $\sum_{i=1}^{n} a_{ij} = 0$, *zero* (resp. *nonpositive*) *row excess* if $\sum_{j=1}^{n} a_{ij} = 0$, and *zero* (resp. *nonpositive*) *excess* if both its row and column excess are zero (resp. nonpositive).

One last general class of matrices to be introduced are the so-called M-matrices. They have been studied very extensively in economics[5], numerical analysis[10], electrical network analysis[6], and there have also been a number of applications to stability analysis in the control literature as well[7, 8, 11–14].

Definition 4. A is said to be an *M-matrix* (resp. *semi M-matrix*) if $a_{ij} \leq 0$ $(i \neq j)$ and if all its principal minors are positive (resp. nonnegative).

There is a very extensive literature on the properties of M-matrices. A good account of these may be found in [5, 15]. Some of their basic properties which will be relevant to this article are:

(I) Let $a_{ij} \leq 0$ $(i \neq j)$. Then the following conditions are equivalent:

(i) A is an M-matrix;
(ii) $-A$ is Hurwitz (i.e. $Re\,\lambda > 0$ for $\lambda \in \sigma(A)$);
(iii) The *leading* principal minors of A are positive;
(iv) There exists a diagonal matrix $\Lambda = \text{diag}(\lambda_i)$ with $\lambda_i > 0$ such that $A\Lambda$ (resp. ΛA) is strictly row (resp. column) dominant.

Note that condition (iii) is particularly interesting since it gives a test (analogous to the Sylvester test for symmetric matrices) for checking whether or not a matrix with nonnegative off-diagonal elements is Hurwitz. Conditions analogous to the above exist for semi M-matrix. However, for the purposes of the present paper one is more interested in a slightly more restricted class, namely those for which $\dot{x} + Ax = 0$ is stable (if A is a semi M-matrix then one can only guarantee that $Re\,\lambda \geq 0$ for $\lambda \in \sigma(A)$ which is just a bit short of stability).

Definition 5. A is said to be *indecomposable* if there does not exist a non-empty subset J of $\{1, 2, \ldots, n\}$ such that $a_{ij} = 0$ $(i \notin J, j \in J)$.

The conditions analogous to (I) are:

(II) Let A be indecomposable with $a_{ij} \leq 0$ $(i \neq j)$. Then the following conditions are equivalent:

(i)′ A is a semi M-matrix;
(ii)′ $\dot{x} + Ax = 0$ is stable;
(iv)′ There exists a diagonal matrix $\Lambda = \text{diag}(\lambda_i)$ with $\lambda_i > 0$ such that $A\Lambda$ (resp. ΛA) is row (resp. column) dominant.

Note that (iv) and (iv)′ imply a connection between dominant and M-matrices obtained by taking $\Lambda = I$.

There is a close connection between M-matrices and nonnegative matrices. Thus is is well known that a nonnegative matrix A has its spectral radius $\rho(A) < 1$ (resp. ≤ 1) iff $I - A$ is an M-matrix (resp. semi M-matrix). A further connection involving matrix exponentials will be stated in Section 4.

3. *Basic inequalities*

In this section certain inequalities are proven involving the matrices introduced in the previous section and certain classes of convex functions:

Definition 6. Let $V: \mathbb{R}^n \to \mathbb{R}$. Then V is said to be *convex* on D (a convex subset of $\mathbb{R}^n$) if $V(\alpha x_1 + (1-\alpha)x_2) \leq \alpha V(x_1) + (1-\alpha)V(x_2)$, $0 \leq \alpha \leq 1$, $x_1, x_2 \in D$; it is said to be *invariant under coordinate permutations if* $V(Px) = V(x)$ for any permutation matrix P, and *invariant under sign reversal of coordinates* if $V(Sx) = V(x)$ for any signature matrix S (i.e. any diagonal matrix S with $S^2 = I$).

If V is sufficiently differentiable, then its convexity is most easily verified from examining its second derivative matrix. If D in Definition 6 is all of $\mathbb{R}^n$ then V will simply be called convex. Finally, recall then any norm is convex.

The folowing proposition gives the basic inequalities which will be used for constructing Lyapunov functions in Section 5:

Proposition 1. (i) *$V(Ax) \leq V(x)$ whenever V is convex, invariant under coordinate permutations, and invariant under sign reversal of the coordinates iff $\sum_i |a_{ij}| \leq 1$ and $\sum_j |a_{ij}| \leq 1$;*

(ii) *$V(Ax) \leq V(x)$ whenever V is convex, invariant under coordinate permutations, and $V(0) \leq V(x)$ iff A is nonnegative, $\sum_i a_{ij} \leq 1$, and $\sum_j a_{ij} \leq 1$;*

(iii) *$V(Ax) \leq V(x)$ whenever V is convex and invariant under coordinate permutations iff A is doubly stochastic.*

Proof. (*if*): only case (i) is proven explicitly; the proofs of (ii) and (iii) are completely analogous. Clearly $|A|$ is substochastic by assumption. Hence there exist subpermutation matrices P_k and $\gamma_k \geq 0$, with $\sum_k \gamma_k = 1$, such that $|A| = \sum_k \gamma_k P_k$. The following string of inequalities then proves the claim

$$\begin{aligned} V(Ax) &= V(|Ax|) \leq V(|A||x|) \\ &= V\Big(\sum_k \gamma_k P_k |x|\Big) \leq \sum_k \gamma_k V(P_k|x|) \\ &\leq \sum_k \gamma_k V(|x|) = V(|x|) = V(x). \end{aligned}$$

■

(*only if*): again only case (i) is considered. Clearly $\|x\|_1 = \sum_i |x_i|$ and $\|x\|_\infty = \max_i |x_i|$ are admissible V-functions. It is well-known[16] that the corresponding induced norms $\|A\|_1 = \max_j \sum_i |a_{ij}|$ and $\|A\|_\infty = \max_i \sum_j |a_{ij}|$, which proves the claim.

■

The classes of functions V considered in Proposition 1 only depend on the matrix A is a very global way. This is, however, not the case for the next proposition since the coefficients λ_i and μ_i of the V functions will in general be different for every A.

Proposition 2. *Assume that $\lambda_i \geq 0$, $\mu_i > 0$, $f: \mathbb{R} \to \mathbb{R}$, and $V(x) = \sum_i \lambda_i f(\mu_i x_i)$. Then $V(Ax) \leq V(x)$ whenever*

(i) *f is convex and $f(\sigma) = f(-\sigma)$ iff $\sum_i \lambda_i \mu_i |a_{ij}| \leq \lambda_j \mu_j$ and $\sum_j \frac{|a_{ij}|}{\mu_j} \leq \frac{1}{\mu_i}$;*

(ii) *f is convex and $f(\sigma) \geq f(0)$ iff A is also nonnegative;*

(iii) *f is convex iff A is nonnegative, $\sum_i \lambda_i \mu_i a_{ij} = \lambda_j \mu_j$, and $\sum_j \frac{a_{ij}}{\mu_j} = \frac{1}{\mu_i}$.*

Proof. for the sake of brevity, only the "if" part of (i) is considered; the "if" parts of (ii) and (iii) are analogous, and the "only if" part of the proposition may be proven as in Proposition 1. Now,

$$\begin{aligned} \sum_i \lambda_i f\Big(\mu_i \sum_j a_{ij} x_j\Big) &\leq \sum_i \lambda_i f\Big(\mu_i \sum_j |a_{ij}||x_j|\Big) \\ &= \sum_i \lambda_i f\Big(\sum_j \frac{|a_{ij}|\mu_i}{\mu_j} \mu_j |x_j|\Big) \\ &\leq \sum_i \sum_j \frac{\lambda_i \mu_i}{\mu_j} |a_{ij}| f(\mu_j |x_j|) \\ &\leq \sum_j \lambda_j f(\mu_j |x_j|) = \sum_j \lambda_j f(\mu_j x_j). \end{aligned}$$

■

Remarks. 1. Proposition 1 leaves the question whether or not the λ_i and μ_i exist unanswered. This may however be resolved. Indeed:

(i) $\rho(|A|) \leq 1 \Leftrightarrow I - |A|$ is an M-matrix $\Leftrightarrow$ there exist λ_i, $\mu_i > 0$ satisfying the conditions of case (i) with strict inequality.
(ii) if A is indecomposable then $\rho(|A|) \leq 1 \Leftrightarrow I - |A|$ is a semi M-matrix $\Leftrightarrow$ there exist λ_i, $\mu_i > 0$ satisfying the conditions of case (i);
(iii) iff $|A|$ is substochastic one may take $\mu_i = 1$ and iff $|A|$ is doubly substochastic one may take $\mu_i = \lambda_i = 1$;
(iv) if A is indecomposable and nonnegative then $\rho(A) = 1 \Leftrightarrow I - A$ is a singular semi M-matrix $\Leftrightarrow$ there exist λ_i, $\mu_i > 0$ satisfying the conditions of case (iii).

2. Let $\|\cdot\|$ denote an arbitrary norm on $\mathbb{R}^n$. Then Proposition 1 shows that for every norm which satisfies $\|x\| = \|Px\| = \|Sx\|$ for all permutation matrices P and signature matrices S, the induced norm $\|A\| \leq 1$ if $\sum_i |a_{ij}| \leq 1$ and $\sum_j |a_{ij}| \leq 1$. If $\|x\| = \|Px\|$ for all permutation matrices P then $\|A\| \leq 1$ if A is nonnegative and satisfies these inequalities.

3. Proposition 2 and Remark 1 show that if $I-|A|$ is an M-matrix or if A is indecomposable and $I-|A|$ is a semi M-matrix then there exist, for each $1 \le p \le \infty$, constants $\alpha_i > 0$ such that the norm induced by the ℓ_p-type norm $\|x\| \triangleq \left(\sum_{i=1}^{n} \alpha_i |x_i|^p\right)^{1/p}$ satisfies $\|A\| \le 1$.

4. Proposition 2 and Remark 1 provide useful generalizations of the known result stating that $\rho(|A|) < 1 \Leftrightarrow I-|A|$ is an M-matrix $\Leftrightarrow$ there exist $\alpha_i > 0$ such that $\|A\| < 1$ with $\|x\| = \sum_i \alpha_i |x_i| \Leftrightarrow$ there exist $\beta_i > 0$ such that $\|A\| < 1$ with $\|x\| = \max_i \beta_i |x_i|$.

4. *Matrix exponentials*

In this section the various classes of matrices are related through their matrix exponentials. The first result is known:

Proposition 3. (i) $\sum_i |(e^{At})_{ij}| \le 1$ *(resp.* < 1) *for all* $t > 0$ *iff* A *is (resp. strictly) row dominant and* $a_{ii} \le 0$;

(ii) $\sum_i |(e^{At})_{ij}| \le 1$ *(resp.* < 1) *for all* $t > 0$ *iff* A *is (resp. strictly) column dominant and* $a_{ii} \le 0$.

Proof. (ii) follows from (i) by transposition. The "if" part of (i) is well-known (see e.g. [16, p. 21]) and the "only if" part follows readily by examining e^{At} for t sufficiently small. ∎

The next proposition considers positive matrices. The results are again either well-known [1, p. 25] or easy to derive from there:

Proposition 4. A *defines a generator of nonnegative type iff* e^{At} *is nonnegative for all* $t > 0$; *it defines a generator of nonnegative type and has zero (resp. nonpositive) row excess iff* e^{At} *is stochastic (resp. substochastic) for all* $t > 0$; *it defines a generator of nonnegative type and has zero (resp. nonpositive) excess iff* e^{At} *is doubly stochastic (resp. substochastic) for all* $t > 0$.

Turning now to M-matrices one obtains the following characterization of their matrix exponentials. The proof is an immediate consequence of Proposition 4 and property (ii) of M-matrices quoted in Section 2:

Proposition 5. e^{At} *is a nonnegative matrix with* $\rho(e^{At}) < 1$ *(resp.* ≤ 1) *for* $t > 0$ *iff* $-A$ *is an* M*-matrix (resp. semi* M*-matrix).*

The last proposition of this section is similar to Proposition 3 but considers the absolute values of the matrix entries:

Proposition 6. $\rho(|e^{At}|) < 1$ *(resp.* ≤ 1) *for* $t > 0$ *iff* $-A_{diag} - |A_{off}|$ *is an* M*-matrix (resp. semi* M*-matrix).*

Proof. (if) by property (iv) of M-matrices quoted in Section 2 there exists $\lambda_i > 0$ and $\epsilon > 0$ such that $\|x\| = V(x) = \sum_i \lambda_i |x_i|$ satisfies $\dot V(x) \le -\epsilon V(x)$ along solutions of $\dot x = Ax$. Thus $\|e^{At}\| < 1$ for $t > 0$ and since for this norm $\|e^{At}\| = \||e^{At}|\|$, the result follows. The case for semi M-matrices may be proven by considering it as the limit of M-matrices.

(*only if*): Considering e^{At} for small positive t yields $\rho(I + A_{diag}t + |A_{off}|t) \le 1$ for t sufficiently small. This implies that $-A_{diag}t - |A_{off}|t$ is an M-matrix for $t > 0$ sufficiently small. Thus $-A_{diag} - |A_{off}|$ must be an M-matrix. The case of semi M-matrices is completely analogous. ∎

Remark finally that there is a close connection between the indecomposability of A and e^{At}. Indeed it is easy to show that if $a_{ij} < 0$ for $i \ne j$ then e^{At} indecomposable for some $t \ne 0 \Leftrightarrow A$ indecomposable $\Leftrightarrow e^{At}$ indecomposable for all $t \ne 0$. Also $|e^{At}|$ indecomposable for all $t > 0 \Leftrightarrow A$ indecomposable.

5. *Lyanpunov functions*

In this section the results of Sections 3 and 4 will be combined to give Lyapunov functions for $\Sigma: \dot x = Ax$. The concept of a Lyapunov function which will be used here does not require V to be sign definite. This is not entirely standard but in keeping with some modern developments is this area. Let $\mathscr{L}$ denote the class of real valued Lipschitz continuous functions on $\mathbb{R}^n$. Thus for any $V \in \mathscr{L}$ and any solution $x(t)$ of Σ, $V(x(t))$ is an absolutely continuous function of t. Let $\dot V(x)]_\Sigma$ denote the time derivative of $V(e^{At}x)$ at $t = 0$. If $V(x)$ is differentiable at x then $\dot V(x) = (\partial/\partial x)V(x)\cdot Ax$ but in any case $\dot V(x(t))]_\Sigma$ exists almost everywhere along solutions of Σ. Recall that a convex function is Lipschitz continuous.

Definition 7. A function $V \in \mathscr{L}$ will be called a *Lyapunov function* for Σ on the subset D of $\mathbb{R}^n$ if $\dot V(x)]_\Sigma \le 0$ for all $x \in D$ where the derivative exists.

There is of course an obvious analogue of the above definition for discrete time systems. Thus $V(x)$ is a Lyapunov function for the discrete time system $x_{k+1} = Ax_k$ if $V(Ax) \le V(x)$. The results of Section 3 immediately give Lyapunov functions for such systems. With the results of Section 4, however, it may be seen that these also lead to Lyapunov functions for Σ. Indeed V is a Lyapunov function for Σ iff $V(e^{At}x) \le V(x)$ for all $t \ge 0$. Thus Proposition 1, 3 and 4 immediately imply the following result:

Theorem 1: *V is a Lyapunov function for Σ on $\mathbb{R}^n$*

(i) *whenever V is convex, invariant under coordinate permutations, and invariant under sign reversal iff A is doubly dominant and* $a_{ii} \le 0$; *or*

(ii) *whenever V is convex, invariant under coordinate permutations and* $V(0) \le V(x)$ *iff A is a generator of nonnegative type and doubly dominant*; *or*

(iii) *whenever V is convex and invariant under coordinate permutations iff A is a generator of nonnegative type and doubly dominant with zero excess.*

The appealing feature of Theorem 1 is that the conditions on V for it to be a Lyapunov function are very easily recognized and verified. It is clear that one can also obtain an analogous theorem derived from Propositions 2, 5 and 6 and the remarks following them:

Theorem 2. *Let* $f: \mathbb{R} \to \mathbb{R}$ *be convex and* $V(x) = \sum_i \lambda_i f(\mu_i x_i)$. *Then there exist* $\mu_i, \lambda_i > 0$ *such that V is a Lyapunov function for Σ on* $\mathbb{R}^n$

(i) *if* $f(\sigma) = f(-\sigma)$ *and* $-A_{diag} - |A_{off}|$ *is an M-matix or an indecomposable semi M-matrix*; *or*

(ii) *if* $f(0) \le f(\sigma)$ *and* $-A$ *is an M-matrix or an indecomposable semi M-matrix*; *or*

(iii) *if $-A$ is an indecomposable singular semi M-matrix.*

Remarks

1. It is of course possible to state the above theorem as an iff condition. Assume that $\mu_i, \lambda_i > 0$ and that V is a Lyapunov function for some f satisfying $\lim_{|\sigma| \to \infty} f(\sigma) = \infty$. Then obviously $\dot x = Ax$ will be stable. If $a_{ij} \ge 0$ for $i \ne j$, $-A$ will then be a semi M-matrix. If V is such that actually asymptotic stability is guaranteed then $-A$ will be an M-matrix. The important point is that M-matrices have "diagonal" type of Lyapunov functions as shown in Theorem 2.

2. Theorem 1 and 2 provide useful generalizations of the fact that $\sum_{i=1}^{n} |x_i|$ is a Lyapunov function iff A is column dominant with $a_{ii} \le 0$ and $\max_i |x_i|$ is a Lyapunov function iff A is row dominant with $a_{ii} \le 0$.

3. $\max_i |x_i|$ is a Lyapunov function iff A is row dominant with $a_{ii} \le 0$; $\max_i (x_i, 0)$ and $-\min_i (x_i, 0)$ are Lyapunov functions iff A is a generator of nonnegative type and row dominant; and $\max_i x_i$ and $-\min_i x_i$ are Lyapunov functions iff A is a generator of nonnegative type and row dominant with zero excess. This shows the *diffusive* character of such systems.

4. An interesting limit case of Theorem 2 is $\max_i |\mu_i x_i|$ in case (i); $\max_i (\mu_i x_i, 0)$ and $-\min_i (\mu_i x_i, 0)$ in case (ii); and $\max_i \mu_i x_i$ and $-\min_i \mu_i x_i$ in case (iii).

5. Theorems 1 and 2 guarantee that under certain conditions $\dot V]_\Sigma \le 0$. Since $\dot V(x)]_\Sigma = (\partial V/\partial x)(x) \cdot Ax$ this shows that the results will not be particularly dependent on

the fact that A is a constant matrix and consequently the methods lead to stability results for nonlinear and/or time-varying systems of the type $\dot{x} = A(x, t)x$. A given nonlinear system $\dot{x} = f(x, t)$ may be written in this form. In fact, $A(x, t) = \int_0^1 (\partial f/\partial x)(\alpha x, t)\, d\alpha$ is always one such matrix[17]. Thus, for example, if the Jacobian matrix $(\partial f/\partial x)(x, t)$ is strictly row (or column) dominant and has negative diagonal elements with some uniformity in x and t then asymptotic stability will follow with $\max_i |x_i|$ or $\sum_i |x_i|$ as a Lyapunov function[7, 8, 17, 18]. Ref. [17] contains in fact a much improved version of this result which requires only that the jth row of $A(x, t)$ be dominated by its diagonal element in the region $|x_j| = \max_i |x_i|$. Šiljak[13] gives a nonlinear version of Proposition 2. The difficulty there is that one somehow has to guarantee that the μ_i or the λ_i are independent of x and t. By suitable bounding the elements of $A(x, t)$ as described in [13] one may in fact achieve such a situation.

6. Another possible avenue of generalization is to state analogous Lyapunov functions for unstable systems. One such result would be the following: assume that $|A_{diag}| - |A_{off}|$ is an M-matrix or an indecomposable semi-M matrix. Then there will exist $\lambda_i, \mu_i > 0$ such that $\sum_i (\operatorname{sgn} a_{ii})\lambda_i f(\mu_i x_i)$ is a Lyapunov function for Σ on $\mathbf{R}^n$ for any convex f with $f(\sigma) = f(-\sigma)$. In here $\operatorname{sgn} \sigma = \sigma/|\sigma|$ for $\sigma \neq 0$ and $\operatorname{sgn} 0 = 0$. Thus if $|A|_{diag} - |A|_{off}$ is an M-matrix, Σ will be unstable iff $a_{ii} > 0$ for some i.

6. *Applications to Markov chains*

It is customary in Markov chains to arrange the probabilities as a row vector $p = (p_1, p_2, \ldots, p_n)$ where p_i equals the probability that the system is in state i. Consider thus the system

$$\mathcal{M} : \dot{p} = pA$$

Definition 8. $\mathcal{M}$ *is said to define a continuous Markov chain* if e^{At} is a stochastic matrix for all $t \geq 0$. This Markov chain is said to have the *equipartition property* if $p = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ defines an equilibrium point of $\mathcal{M}$.

Thus by Proposition 4 $\mathcal{M}$ defines a Markov chain if and only if A is a generator of positive type with zero row excess. Clearly it has the equipartition property iff e^{At} is doubly stochastic for all $t \geq 0$, i.e. if and only if A is a generator of positive type with zero excess. Let $\mathbf{R}^n$ denote $\{(p_1, p_2, \ldots, p_n), p_i \geq 0\}$. Clearly any Markov chain leaves the convex set $P = \{p \in \mathbf{R}_+^n | \sum_i p_i = 1\}$ invariant, and thus the solution vector $p(t)$ has an obvious probabilistic interpretation. Moreover there is always at least one equilibrium point of $\mathcal{M}$ in P. With the equipartition property, $\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ defines such a point. Moreover if A is indecomposable then there is precisely one asymptotically stable equilibrium state $\pi \in P$ and $\pi_i > 0$.

Theorems 1 and 2 imply the following result:

Theorem 3. *Assume that $\mathcal{M}$ defines a Markov chain. Let $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n) \in P$ be an equilibrium point of $\mathcal{M}$ with $\bar{p}_i > 0$. If f is a convex function on P then $V(p) = \sum_{i=1}^n \bar{p}_i f(p_i/\bar{p}_i)$ is a Lyapunov function for $\mathcal{M}$ on P. If $\mathcal{M}$ has the equipartition property, then any convex function V defined on P which is invariant under coordinate permutations is a Lyapunov function for $\mathcal{M}$ on P.*

These results lead to some special Lyapunov functions which have an interesting information theoretic interpretation. Consider the following standard definition:

Definition 9. Let $(p_1^{(j)}, p_2^{(j)}, \ldots, p_n^{(j)})(j = H_1, H_2)$ be two probability distributions with $p_i^{(j)} > 0$. Then the *average weight of evidence in favor of H_1 against H_2 given H_1* is defined by

$$W(H_1/H_2) = \sum_{i=1}^n p_i^{(H_1)} \log \frac{p_i^{(H_1)}}{p_i^{(H_2)}}.$$

The *divergence between the hypotheses H_1 and H_2* is defined as

$$J(1, 2) = W(H_1/H_2) + W(H_1/H_2).$$

The notions of weight of evidence and divergence have important applications in statistics and in information theory. They have been extensively studied, for example in [19, 20], where further references may be found.

Consider now a continuous Markov chain $\mathcal{M}$ which has an equilibrium point $\bar{p} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_n)$ with $\bar{p}_i > 0$. Assume an initial condition $p(0) \in P$. Let H_T denote the hypothesis "$t = T$" and let H_∞ denote the hypothesis "$t = \infty$", i.e. $p_i^{(H_T)} = p_i(T)$ and $p_i^{(H_\infty)} = \bar{p}_i$. (Note that $\bar{p}$ need *not* be $\lim_{t\to\infty} p(t)$ although it could be, and it is convenient to think of it as such). Theorem 3 implies an interesting and intuitively appealing convergence of the weight of evidence and the divergence between H_T and H_∞.

Theorem 4. *The average weight of evidence in favor of H_T against H_∞ given H_T, the average weight of evidence in favor of H_∞ against H_T given H_∞, and the divergence between the hypotheses H_T and H_∞ are all nonincreasing functions of T.*

Proof. Choose in Theorem 3, $f(\sigma) = \sigma \log \sigma$, $f(\sigma) = -\log \sigma$, and $f(\sigma) = (\sigma - 1) \log \sigma$, $0 \leq \alpha \leq 1$, respectively. ∎

The above theorem adds a new information theoretic quantity which is nonincreasing during the evolution of a Markov chain. It shares this property with the *mutual information* between the input and the output and the *channel capacity* of the channel defined by the Markov chain transition from $t = 0$ (the channel input) to $t = T$ (the channel output).

Remarks.

1. Theorem 3 is of course also valid for discrete Markov chains because the proofs, which are based on Section 3, do not exploit the continuous nature of the chain.

2. The results may be of interest in statistical mechanics where they constitute considerable generalizations of known results[21, 22] in this area.

3. For Markov chains which also have the equipartition property (and only for those!) the *entropy of the chain* at time T defined by

$$H(T) = -\sum_{i=1}^n p_i(T) \log p_i(T)$$

is nondecreasing with T. This follows from Theorem 3. In fact, Theorem 3 also implies that the generalized entropy functions introduced by Arimoto in [23] are nondecreasing with T.

7. *Conclusions*

In this paper it has been shown how one may exploit dominance conditions in order to obtain nonquadratic Lyapunov functions for linear systems. The appealing part of the results lies in the fact that the conditions on the systems and on the Lyapunov functions turn out to be rather simple and easy to verify. It would be interesting to attempt generalizations to systems described in input–output form. This has the potential of leading to an input–output theory of diffusion, absorption and other physically interesting qualitative assumptions.

References

[1] R. W. Brockett: *Finite Dimensional Linear Systems*. Wiley, New York (1970).
[2] V. M. Popov: *Hyperstability of Control Systems*. Springer, Berlin (1973).
[3] R. E. Kalman: Lyapunov functions for the problem of Lur'e in automatic control. *Proc. Nat. Acad. Sci. U.S.A.* 49, 201–205 (1963).
[4] J. C. Willems: Dissipative dynamical systems. Part I: general theory. *Arch Rat. Mech. Anal.* 45, 321–351 (1972).
[5] H. Nikaido: *Convex Structures and Economic Theory*. Academic Press, New York (1968).
[6] See papers by I. W. Sandberg and A. N. Willson, Jr.

in *Nonlinear Networks: Theory and Analysis*. (Ed.: A. N. WILLSON, JR.), IEEE Press, New York (1975).
[7] H. H. ROSENBROCK: Lyapunov functions with applications to some nonlinear physical systems. *Automatica* **1**, 31-53 (1963).
[8] H. H. ROSENBROCK: A lyapunov function for some naturally-occurring linear homogeneous time-dependent equations. *Automatica* **1**, 97-109 (1963).
[9] L. MIRSKY: Results and problems in the theory of doubly-stochastic matrices. *Z. Wahrscheinlichkeitstheorie* **1**, 315-334 (1963).
[10] A. OSTROWSKI: On some metrical properties of operator matrices and matrices partitioned into blocks. *J. Math. Anal. Appl.* **2**, 161-209 (1961).
[11] P. A. COOK: On the stability of interconnected systems. *Int. J. Control* **20**, 407-415 (1974).
[12] M. ARAKI and B. KONDO: Stability and transient behavior of composite nonlinear systems. *IEEE Trans. Aut. Control* **AC-17**, 537-541 (1972).
[13] D. D. ŠILJAK: Connective stability of compatitive equilibrium. *Automatica* **11** (1975).
[14] H. H. ROSENBROCK: *Computer-Aided Control System Design*. Academic Press, London (1974).
[15] M. ARAKI: *M*-Matrices (Matrices with nonpositive off-diagonal elements and positive principal minors). Publication 74/19, Dept. of Comp. and Control, Imperial College, London (1974).
[16] C. A. DESOER and M. VIDYASAGAR: *Feedback Systems: Input–Output Properties*, Academic Press, New York (1975).
[17] J. P. LASALLE and E. N. ONWUCHEKWA: An invariance principle for vector lyapunov functions. Manuscript Division of Appl. Math., Brown Un., Providence, R.I. (1974).
[18] C. KAHANE: Stability of solutions of linear systems with dominant main diagonal. *Proc. Am. Math. Soc.* **33**, 69-71 (1972).
[19] S. KULLBACK: *Information Theory and Statistics*. Wiley, New York (1959).
[20] D. B. OSTEYEE and I. J. GOOD: *Information, Weight of Evidence, the Singularity between Probability Measures and Signal Detection*. Springer Verlag, Berlin. Lecture Notes in Mathematics, Number 376 (1974).
[21] O. PENROSE: *Foundations of Statistical Mechanics*. Pergamon Press, Oxford (1970).
[22] E. RUCH: The diagram lattice as structural principle. *Theoret. Chim. Acta* **38**, 167-183 (1975).
[23] S. ARIMOTO: Information—theoretical considerations on estimation problems. *Inform. Control* **19**, 181-194 (1971).