JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 33, 574–604 (1971)

Optimization of Nonlinear Characteristics*

M. K. INAN AND C. A. DESOER

Department of Electrical Engineering and Computer Sciences,
Electronics Research Laboratory,
University of California, Berkeley, California 94720

Submitted by L. A. Zadeh

1. INTRODUCTION

The problem of minimizing a performance criterion of a dynamical system with respect to the forcing function $u(t)$ of the nonlinear differential equation is well known in optimization theory. Theoretical answers to this problem are given by the Maximum Principle of Pontryagin [5] or methods of Variational Calculus [4].

Another optimization problem is that of selecting the best characteristic of a device to be used within a dynamical system. For example consider a device whose input is $cx(t)$, a linear combination of the state variables at time t, and whose output is $N(cx(t))$ where $N(\cdot)$ maps R into R. The problem is, given some suitable restrictions on $N(\cdot)$ and a performance criterion, find the characteristic $N^*(\cdot)$ which minimizes the criterion. With respect to the problem mentioned in the previous paragraph the present problem is equivalent to finding the optimal singleloop feedback law—not necessarily linear—which realizes the input $u(t)$ by state feedback, namely $u(t) = N(cx(t))$.

The problem posed above is of considerable practical importance. Some possible applications are given below.

A. *Stability.*

A numerical investigation of the stability of nonlinear feedback systems requires the computation of the nonlinearity in the feedback loop that makes the system most and least stable on a time interval chosen large enough to detect the behaviour of the system. A knowledge of the shape of the nonlinearity that makes a system, which is on the boundary of stability, unstable may give better insight in forming new stability criteria.

* Research sponsored by the National Science Foundation under Grant GK-2277 and National Aeronautics and Space Administration under Grant NGL 05 003-016 (Sup 6). The results of this paper were presented at the seventh Allerton Conference, October 10, 1969.

B. *Design.*

The field of system design is full of such applications. For example in circuits it is often a desirable objective to know the characteristics of a nonlinear resistor or amplifier that optimizes an index such as efficiency or power output. It may be cheaper and easier to build a nonlinear resistor or amplifier of a given characteristics than to apply a signal $u(t)$ to optimize the system.

C. *Sensitivity.*

In most systems the characteristics of the nonlinearities vary slowly with time, therefore it is important to know the sensitivity of the performance of the system with respect to certain parameters of the nonlinearity such as its slope at a given point etc. The techniques developed in this paper are helpful in determining such sensitivity coefficients.

The problem of optimizing nonlinear characteristics has been a neglected area of research. A simple case has been considered by Willems [8] in which the system had no dynamics. He considered a single-input single-output nonlinearity subject to a fixed input on a fixed time interval. The objective was to optimize an index that depended on the input output pair on the fixed time interval. A more experimental approach was taken by Soudack [7] for generating the sensitivity coefficients of a system with respect to the slopes of a piecewise affine nonlinearity using an analog computer.

In this paper we derive necessary conditions of optimality for a single-input single-output, memoryless, time-invariant nonlinearity in a very general framework covering finite-dimensional dynamical systems. In Section 2 the problem is formulated. In Section 3 the main result is stated as a theorem and is proved. In Section 4 it is shown that the hypothesis of Theorem 3.1 is satisfied for time-invariant nonlinear feedback systems if the linear time-invariant system in the forward loop has a completely observable representation. In Section 5 the results are generalized for non-linearities with sector constraints. In Section 6 some extensions of the results are pointed out and a similar version of Theorem 3.1 is given for discrete dynamical systems.

2. Formulation of the Problem

A. *System Equations and Properties.*

The system to be considered is described by the following non-linear differential equation.

$$\dot{x}(t) = f(x(t), N(cx(t))), \tag{1}$$

where

$$x(t) = \text{column } n\text{-vector} \qquad \text{for} \qquad t \in [0, T], \qquad T > 0 \textit{ fixed.}$$
$$x(0) = x_0\,, \qquad \textit{fixed,}$$
$$c = \text{constant row } n\text{-vector.}$$

The function $f(\cdot, \cdot)$, maps $R^n \times R$ into R^n, is continuously differentiable and satisfies the following Lipschitz condition $\exists$ constants K_1, K_2 such that

$$|f(x_1, u_1) - f(x_2, u_2)| \leqslant K_1 |x_1 - x_2| + K_2 |u_1 - u_2|\,, \qquad \forall (x_1, u_1), \quad (x_2, u_2) \in R^n \times R. \tag{2}$$

The class of admissible nonlinearities $\mathcal{N}_R$, is described by the following definition.

DEFINITION. $\mathcal{N}_R$ is a family of functions, such that $N \in \mathcal{N}_R$ iff

[D1] $N : R \to R$.

[D2] $N(0) = 0$.

[D3] There is a constant $S > 0$, not depending on N, such that

$$|N(y_1) - N(y_2)| \leqslant S |y_1 - y_2|\,, \qquad \forall y_1, y_2 \in R.$$

[D4] $dN(y)/dy$ is a piecewise continuous[1] function defined almost everywhere on R.

The definition of derivative from the right and the derivative from the left of a real valued function $g(\cdot)$ on R is given below.

$$\frac{d^+g(y)}{dy} \triangleq \lim_{\substack{\gamma > 0 \\ \gamma \to 0}} \frac{N(y + \gamma) - N(y)}{\gamma}, \tag{3a}$$

$$\frac{d^-g(y)}{dy} \triangleq \lim_{\substack{\gamma > 0 \\ \gamma \to 0}} \frac{N(y) - N(y - \gamma)}{\gamma}. \tag{3b}$$

The lemma below describes a property of $\mathcal{N}_R$ which is used later on. The proof is straightforward and is left to the reader.

LEMMA 2.1. *If $N \in \mathcal{N}_R$ then $d^+N(y)/dy$ and $d^-N(y)/dy$ exist for all y in R and are bounded between $-S$ and $+S$.*

Remark. For each $N \in \mathcal{N}_R$ there exists a unique solution $x_N(t)$ to (1) on $[0, T]$ [1, Chapter 1]. Furthermore $x_N(\cdot) \in C^{(1)}$ by the definition of f and $\mathcal{N}_R$.

[1] A function mapping R into R is said to be piecewise continuous iff on each finite interval it has finite number of discontinuities and it is regulated ([2], page 139).

The following lemma introduces simplifications in formulating the problem.

LEMMA 2.2. *There exists constants a and b such that*

$$a < cx_N(t) < b \qquad \forall t \in [0, T], \qquad \forall N \in \mathcal{N}_R . \tag{4}$$

Lemma 2.2 follows directly from the finiteness of $[0, T]$, the Lipschitz condition (2), the slope condition [D3], [D2] and the Bellman-Cronwall inequality.

We define $\mathcal{N}_{[a,b]}$ by restricting the domain of each member of $\mathcal{N}_R$ to the closed interval $[a, b]$. Then, in view of Lemma 2.2, Eq. (1) is well defined for all N in $\mathcal{N}_{[a,b]}$.

B. *The Minimization Problem.*

The functional P is defined as follows:

$$P(N) \triangleq \int_0^T h(x_N(t))\, dt, \qquad N \in \mathcal{N}_{[a,b]}, \tag{5}$$

where

$$h : R^n \to R, \qquad h(\cdot) \in C^1.$$

The basic problem is to minimize $P(N)$ on $\mathcal{N}_{[a,b]}$:

$$\operatorname*{Min}_{N \in \mathcal{N}_{[a,b]}} P(N). \tag{6}$$

By putting sup norm on $\mathcal{N}_{[a,b]}$ it becomes a subset of the Banach Space $C_{[a,b]}$([3]). This construction enables us to talk about a *local* minimum since the set over which we minimize has a relative norm topology. The norm on $\mathcal{N}_{[a,b]}$ is denoted by:

$$| N |_\infty \triangleq \max_{y \in [a,b]} | N(y)| . \tag{7}$$

Notation.

$$\mathcal{N} \triangleq \mathcal{N}_{[a,b]}, \qquad \text{and} \qquad \| x(\cdot)\| = \sup_{t \in [0,T]} | x(t)| .$$

3. MAIN RESULT

THEOREM 3.1. *Suppose N^* furnishes a local minimum for P for the basic problem given by* (6). *Assume that*

$$\mathcal{H} \triangleq \{t \in [0, T]; c\dot{x}^*(t) = 0\} \tag{8}$$

is a finite set given by,

$$\mathscr{H} = \{t_j\}_{j=1}^{m}, \qquad t_j \in [0, T], \qquad j = 1,..., m, \tag{9}$$

where $x^*(t)$ is the optimum trajectory of (1) *corresponding to N^*.*

Under these conditions for all z (except possibly a finite set) in $[a, b]$ the following relations hold

$$\textit{For } z \geqslant 0, \qquad \frac{d^+N^*(z)}{dz} = -S \operatorname{sgn}\left[\int_{I^+(z)} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt\right]; \tag{10a}$$

$$\textit{For } z \leqslant 0, \qquad \frac{d^-N^*(z)}{dz} = +S \operatorname{sgn}\left[\int_{I^-(z)} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt\right]; \tag{10b}$$

where,

$$I^+(z) \triangleq \{t \in [0, T];\ cx^*(t) > z\}, \tag{11a}$$

$$I^-(z) \triangleq \{t \in [0, T];\ cx^*(t) < z\}, \tag{11b}$$

and

$$\left[\frac{\partial f}{\partial N}\right]_{\text{opt}} \triangleq \frac{\partial f(x, u)}{\partial u}\bigg|_{(x,u)=(x^*(t),N^*(cx^*(t))}. \tag{12}$$

$\lambda^*(t)$ solves the adjoint equation[2] *given by*

$$\dot{\lambda}^*(t) = -\left[\frac{df(x, N^*(cx))}{dx}\right]^T_{x=x^*(t)} \lambda^*(t) - \left[\frac{dh(x)}{dx}\right]^T_{x=x^*(t)}, \tag{13}$$

with

$$\lambda^*(T) = 0. \tag{14}$$

Remark. Theorem 3.1 asserts that unless the argument of sgn function is zero the slope of the optimum nonlinearity is on the boundary of the constraint set; i.e. $+S$ or $-S$. Therefore the result is analogous to bang-bang type control problems. The only assumption that may look unreasonable is the one given by (9). However in Section 4 the existence of a large class of feedback systems satisfying (9) is exhibited.

Preliminaries. We investigate the set $I^+(z)$ given by (11a) in more detail. To be definite we assume that (see Fig. 1)

(i) $$z \geqslant 0 \tag{15}$$

(ii) $$cx^*(0) = cx_0 < z, \qquad cx^*(T) < z. \tag{16}$$

[2] In order for (13) to be well defined $dN^*(y)/dy\,|_{y=cx^*(t)}$ must be defined for almost all t in $[0, T]$ which is insured by (9) and [D4].

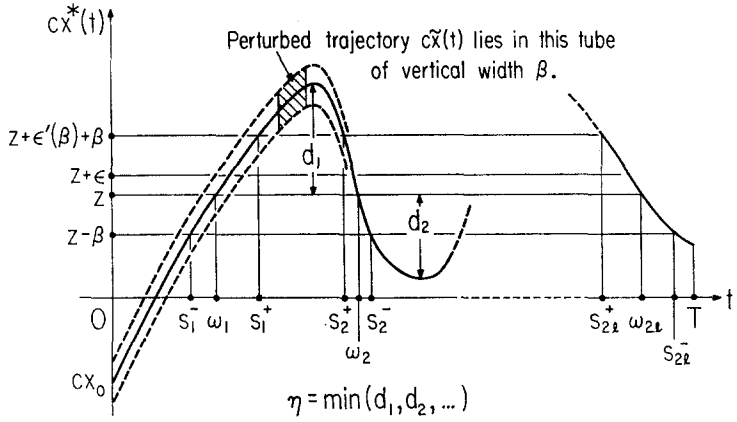

FIG. 1. The plot of t vs $cx^*(t)$ relating the numbers β, $\epsilon'(\beta)$, ω_i, $s_i^{\pm}$.

The remaining seven cases are substantially the same, so they will not be considered in detail. Choose z such that it satisfies (15) and (16). Furthermore assume that z has the property that

$$cx^*(t_j) \neq z, \qquad t_j \in \mathscr{H}, \qquad j = 1,..., m. \tag{17}$$

Clearly all except a finite number of z's in (a, b) satisfy (17). Let

$$\Omega_z \triangleq \{\omega \in [0, T]; cx^*(\omega) = z\}. \tag{18}$$

We claim that Ω_z is a finite set. For if it is not, then by compactness of $[0, T]$ Ω_z has a limit point ω' in $[0, T]$. Furthermore by continuity of $cx^*(\cdot)$, $\omega' \in \Omega_z$. So we have

$$c\dot{x}^*(\omega') \neq 0, \qquad \text{and} \qquad cx^*(\omega') = z.$$

So there is a neighbourhood of ω' such that $cx^*(\omega) \neq z$ for all ω in this neighbourhood except $\omega = \omega'$. This contradicts that ω' is a limit point of Ω_z.

The set Ω_z can therefore be written as

$$\Omega_z = \{\omega_i\}_{i=1}^{2\ell}. \tag{19}$$

The fact that Ω_z has even number of elements is easily deduced from (16) and (17). Now, from (11a) and (19)

$$I^+(z) = \bigcup_{i=1}^{\ell} (\omega_{2i-1}, \omega_{2i}), \tag{20}$$

where we have subscripted the elements of Ω_z by the rule

$$\omega_j > \omega_i \qquad \text{iff} \qquad j > i \qquad \forall i, j = 1,..., 2\ell. \tag{21}$$

Since $I^+(z)$ is a union of disjoint intervals we obtain

$$\int_{I^+(z)} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt = \sum_{i=1}^{\ell} \int_{\omega_{2i-1}}^{\omega_{2i}} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt. \tag{22}$$

We now construct a perturbed nonlinearity as follows (see Fig. 2)

$$\tilde{N}_{\epsilon,k,z}(y) \triangleq \begin{cases} N^*(y), & a \leqslant y \leqslant z \\ N^*(z) + k(y - z), & z < y < z + \epsilon, \\ N^*(y) + k\epsilon - (N^*(z + \epsilon) - N^*(z)), & z + \epsilon < y \leqslant b. \end{cases}$$

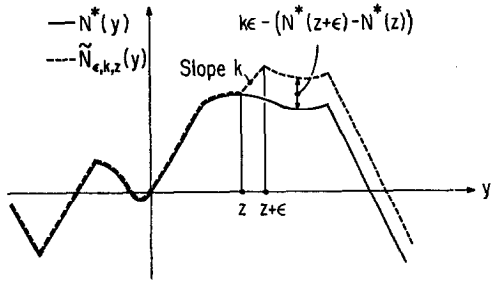

FIG. 2. The perturbed nonlinearity $\tilde{N}_{\epsilon,k,z}$ used in the proof of Theorem 3.1.

Note that

$$\tilde{N}_{\epsilon,k,z} \in \mathcal{N}, \qquad \forall \epsilon > 0, \qquad \forall z \in [0, b], \qquad \forall k \ni |k| \leqslant S, \tag{23}$$

and

$$\frac{|N^* - N_{\epsilon,k,z}|_\infty}{\epsilon} \leqslant 2S, \qquad \forall \epsilon > 0. \tag{24}$$

For convenience we drop the subscripts of $\tilde{N}_{\epsilon,k,z}$ and use the following notation

$$\tilde{x}(t) \triangleq x_{\tilde{N}}(t), \qquad \tilde{y}(t) \triangleq c\tilde{x}(t), \qquad y^*(t) \triangleq cx^*(t).$$

The following lemmas will be repeatedly used in the proof of Theorem 3.1. The proofs of these lemmas are given in the Appendix.

LEMMA 3.1. *For any $\beta > 0$ sufficiently small $\exists \epsilon'(\beta) > 0$ with the property that* (see Fig. 1):

(a) *To each* $\omega_{2i}[\omega_{2i-1}]$ *in* Ω_z *there corresponds an open interval* (s_{2i}^+, s_{2i}^-) $[(s_{2i-1}^-, s_{2i-1}^+)]$ *such that*

$$y^*(s_{2i-1}^-) = y^*(s_{2i}^-) = z - \beta \tag{25a}$$

$$y^*(s_{2i-1}^+) = y^*(s_{2i}^+) = z + \beta + \epsilon'(\beta). \tag{25b}$$

Furthermore

$$\lim_{\beta\to 0} s_{2i}^+ = \lim_{\beta\to 0} s_{2i}^- = \omega_{2i}\,, \qquad [\lim_{\beta\to 0} s_{2i-1}^- = \lim_{\beta\to 0} s_{2i-1}^+ = \omega_{2i-1}], \tag{26a}$$

where

$$\omega_{2i} \in (s_{2i}^+, s_{2i}^-) \qquad [\omega_{2i-1} \in (s_{2i-1}^-, s_{2i-1}^+)], \qquad \forall i = 1, 2, \ldots, \ell. \tag{26b}$$

The s_i*'s are ordered as follows:*

$$0 < s_1^- < s_1^+ < s_2^+ < s_2^- < s_3^- < \cdots < s_{2\ell}^+ < s_{2\ell}^- < T. \tag{27}$$

(b) *For* $\epsilon < \epsilon'(\beta)$ *the following relations hold*

$$\tilde{y}(t) > z + \epsilon, \qquad \forall t \in \bigcup_{i=1}^{\ell} (s_{2i-1}^+, s_{2i}^+), \tag{28a}$$

$$\tilde{y}(t) < z, \qquad \forall t \in [0, s_1^-) \cup (s_{2\ell}^-, T] \cup \left\{\bigcup_{i=1}^{\ell-1} (s_{2i}^-, s_{2i+1}^-)\right\}, \tag{28b}$$

where $\tilde{y}(t)$ *corresponds to* $\tilde{N}_{\epsilon,k,z}$.

LEMMA 3.2. *For all* $i = 1, \ldots, \ell$

(a) $$x^*(s_{2i-1}^+) - \tilde{x}(s_{2i-1}^+) = x^*(s_{2i-1}^-) - \tilde{x}(s_{2i-1}^-) + o(\epsilon, \beta) \tag{29a}$$

(b) $$x^*(s_{2i}^-) - \tilde{x}(s_{2i}^-) = x^*(s_{2i}^+) - \tilde{x}(s_{2i}^+) + o(\epsilon, \beta) \tag{29b}$$

where $o(\epsilon, \beta)$ *has the property*

$$\lim_{\beta\to 0} \left(\overline{\lim_{\substack{\epsilon\to 0 \\ \epsilon>0}}} \frac{|\, o(\epsilon, \beta)|}{\epsilon} \right) = 0. \tag{30}$$

LEMMA 3.3. *Suppose* $x^*(t')$ *is perturbed by* $\delta x'$ *and* $N^*(\cdot)$ *by* γ,[3] *where* $t' \in [0, T)$ *and* γ *is a real number. Then the corresponding trajectory for* $t \in [t', T]$, *denoted by* $\bar{x}(t)$, *is given by*

$$\bar{x}(t) = x^*(t) + \Phi(t, t')\,\delta x' + \gamma \int_{t'}^{t} \Phi(t, \tau) \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau + o(|\,\delta x'\,| + |\,\gamma\,|, t), \tag{31}$$

[3] γ denotes the function in $C[a, b]$ which is identically equal to a constant.

where

$$\lim_{(|\delta x'|+|\gamma|)\to 0} \frac{|\, o(|\,\delta x'\,|+|\,\gamma\,|\,, t)|}{|\,\delta x'\,|+|\,\gamma\,|} = 0 \tag{32}$$

uniformly in $t \in [t', T]$.

$\Phi(\cdot, \cdot)$ is the state transition matrix of the variational equation given by

$$\dot{v}(t) = \left[\frac{df(x, N^*(cx))}{dx}\right]_{x=x^*(t)} v(t). \tag{33}$$

Proof of Theorem 3.1. The proof is by contradiction so we assume that (10a) is not satisfied and z satisfies (15), (16) and (17).

Consider the augmented system

$$\dot{\hat{x}} = \hat{f}(\hat{x}, N(\hat{c}\hat{x})), \tag{34}$$

where

$$\hat{f}(\hat{x}, N(\hat{c}\hat{x})) = \begin{bmatrix} f(x, N(cx)) \\ h(x) \end{bmatrix}, \tag{35a}$$

$$\hat{c} = [c\ 0], \qquad \text{constant row } (n+1)\text{–vector}, \tag{35b}$$

$$\hat{x}(t) = \begin{bmatrix} x(t) \\ x^0(t) \end{bmatrix}, \qquad \text{column } (n+1)\text{–vector for } t \in [0, T], \tag{35c}$$

and

$$\hat{x}(0) = \begin{bmatrix} x_0 \\ 0 \end{bmatrix}. \tag{35d}$$

It follows by the definitions above that

$$x^0(T) = P(N). \tag{36}$$

It can easily be shown that the results of Lemma 3.1 to Lemma 3.3 also hold for the augmented system.

We now prove the following relations

$$\tilde{\hat{x}}(s_{2i}^+) = \hat{x}^*(s_{2i}^+) + \epsilon \left(k - \frac{d^+N^*(z)}{dz}\right) \sum_{j=1}^{i} \left\{ \int_{\omega_{2j-1}}^{\omega_{2j}} \hat{\Phi}(s_{2i}^+, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau \right\} + o(\epsilon, \beta), \qquad i = 1,..., \ell; \tag{37a}$$

$$\tilde{\hat{x}}(s_{2i+1}^-) = \hat{x}^*(s_{2i+1}^-) + \left(k - \frac{d^+N^*(z)}{dz}\right) \sum_{j=1}^{i} \left\{ \int_{\omega_{2j-1}}^{\omega_{2j}} \hat{\Phi}(s_{2i+1}^-, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau \right\} + o(\epsilon, \beta), \qquad i = 1,..., \ell - 1. \tag{37b}$$

where, $o(\epsilon, \beta)$ has the property defined by (30) and $\hat{\Phi}(\cdot, \cdot)$ is the state transition matrix of the variational equation of (34) around the optimal trajectory $\hat{x}^*(t)$. ϵ is chosen small enough to satisfy conditions of Lemma 3.1 to Lemma 3.3 whenever they are applied.

We first prove (37a) and (37b) for $i = 1$. Using Lemma 3.1(b)

$$\tilde{\hat{y}}(t) \triangleq \hat{c}\tilde{\hat{x}}(t) < z \qquad \text{for } t \in [0, s_1^-).$$

So using definition of $\tilde{N}$ we have

$$\hat{x}^*(s_1^-) = \tilde{\hat{x}}(s_1^-).$$

By Lemma 3.2(a)

$$\tilde{\hat{x}}(s_1^+) = \hat{x}^*(s_1^+) + o(\epsilon, \beta). \tag{38}$$

Again using Lemma 3.1(b)

$$\tilde{\hat{y}}(t) > z + \epsilon, \qquad \text{for } t \in (s_1^+, s_2^+). \tag{39}$$

Using definition of $\tilde{N}$ we apply Lemma 3.3 with

$$t' = s_1^+, \qquad \delta x' = o(\epsilon, \beta), \qquad \gamma = k\epsilon - (N^*(z + \epsilon) - N^*(z)),$$

$$\begin{aligned}\tilde{\hat{x}}(s_2^+) = \hat{x}^*(s_2^+) &+ \hat{\Phi}(s_2^+, s_1^+)\, o(\epsilon, \beta) \\ &+ (k\epsilon - (N^*(z + \epsilon) - N^*(z))) \int_{s_1^+}^{s_2^+} \hat{\Phi}(s_2^+, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau \\ &+ o(|\, \delta x' \,| + |\, \gamma \,|, s_2^+).\end{aligned} \tag{40}$$

By continuity of the integral with respect to its end points, using Lemma 3.1(a), and Lemma 2.1 (40) reduces to (37a) for $i = 1$.

Again using Lemma 3.2(b) and then Lemma 3.3 with

$$t' = s_2^-, \qquad \delta x' = \epsilon \left(k - \frac{d^+N^*(z)}{dz}\right) \int_{\omega_1}^{\omega_2} \hat{\Phi}(s_2^+, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau + o(\epsilon, \beta),$$

$$\gamma = 0,$$

$$\begin{aligned}\tilde{\hat{x}}(s_3^-) = \hat{x}^*(s_3^-) &+ \hat{\Phi}(s_3^-, s_2^-)\, \epsilon \left(k - \frac{d^+N^*(z)}{dz}\right) \int_{\omega_1}^{\omega_2} \hat{\Phi}(s_2^+, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau \\ &+ o(\epsilon, \beta).\end{aligned} \tag{41}$$

By using continuity arguments and Lemma 3.1(a) it can be shown that the remainder terms after replacing $\hat{\Phi}(s_3^-, s_2^-)\, \hat{\Phi}(s_2^+, \tau)$ by $\hat{\Phi}(s_3^-, \tau)$ of $o(\epsilon, \beta)$ type therefore (41) reduces to (37b) for $i = 1$.

A routine induction procedure proves (37a) and (37b). Setting $i = \ell$ in (37a) and using Lemma 3.2(b), Lemma 3.1(b) and Lemma 3.3 we obtain the following equation.

$$\tilde{\hat{x}}(T) = \hat{x}^*(T) + \epsilon \left(k - \frac{d^+N^*(z)}{dz}\right) \sum_{j=1}^{\ell} \left\{ \int_{\omega_{2j-1}}^{\omega_{2j}} \hat{\Phi}(T, \tau) \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} d\tau \right\} + o(\epsilon, \beta). \tag{42}$$

We now let $\hat{\lambda}^*(t)$ be the solution of the augmented adjoint equation given by

$$\dot{\hat{\lambda}}^*(t) = -\left[\frac{d\hat{f}(\hat{x}, N^*(\hat{c}\hat{x}))}{d\hat{x}}\right]^T_{\hat{x}=\hat{x}^*(t)} \hat{\lambda}^*(t), \tag{43}$$

where

$$\hat{\lambda}^*(T) = \begin{bmatrix} \lambda^*(T) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{44}$$

Then by standard results in optimization [5] we obtain the following relation

$$\hat{\lambda}^*(t) = \begin{bmatrix} \lambda^*(t) \\ 1 \end{bmatrix}, \qquad \forall t \in [0, T], \tag{45}$$

where $\lambda^*(t)$ is given by (13) and (14). Using (36), (44) and (45) we obtain

$$(\hat{\lambda}^*(T))^T (\tilde{\hat{x}}(T) - \hat{x}^*(T)) = \tilde{x}^0(T) - x^{0*}(T) = P(\tilde{N}) - P(N^*), \tag{46}$$

$$(\hat{\lambda}^*(T))^T \hat{\Phi}(T, \tau) = (\hat{\lambda}^*(\tau))^T, \tag{47}$$

and

$$(\hat{\lambda}^*(\tau))^T \left[\frac{\partial \hat{f}}{\partial N}\right]_{\text{opt}} = (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}}. \tag{48}$$

Taking inner product of both sides of (42) with $\hat{\lambda}^*(T)$ and using (46), (47), (48) and (22)

$$P(\tilde{N}) - P(N^*) = \epsilon \left(k - \frac{d^+N^*(z)}{dz}\right) \int_{I^+(z)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau + o(\epsilon, \beta). \tag{49}$$

For ϵ and β small enough we have that

$$\operatorname{sgn}[P(\tilde{N}) - P(N^*)] = \operatorname{sgn}\left[\left(k - \frac{d^+N^*(z)}{dz}\right) \int_{I^+(z)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau\right]. \tag{50}$$

Since k is an arbitrary number between $+S$ and $-S$ if (10a) is not satisfied we obtain

$$\operatorname{sgn}[P(\tilde{N}) - P(N^*)] < 0, \tag{51}$$

which contradicts optimality of N^*.

This completes the proof of Theorem 3.1.

4. Application

In this section we specialize to the nonlinear feedback-system given by Fig. 3.

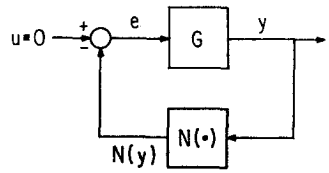

Fig. 3. The nonlinear feedback system block diagram used in section 4.

G is a linear dynamical system described by the following equations:

$$\dot{x} = Ax + be, \tag{52a}$$

$$y = cx, \tag{52b}$$

where

$$A = \text{constant } n \times n \text{ matrix},$$

$$b = \text{constant column } n\text{-vector},$$

$$c = \text{constant row } n\text{-vector},$$

N is an element of $\mathcal{N}$, as defined in Section 2.

The equations describing the dynamics of the overall feedback system can be written as

$$\dot{x} = Ax - bN(cx), \tag{53a}$$

and

$$y = cx. \tag{53b}$$

The problem is the basic problem given by (6) where (1) is specialized to the Eq. (53a). It is easily seen that rightside of (53a) satisfies the conditions on f given in Section 2.

We now state the main result of this section.

THEOREM 4.1. *Suppose N^* furnishes a local minimum for P with $N^* \in \mathcal{N}$. Assume that G is completely observable, then either the assumption given by* (9) *of Theorem 3.1 is satisfied or $x^*(t)$ is a trivial trajectory of* (53a), *in other words,*

$$x^*(t) = x_0, \qquad \forall t \in [0, T], \tag{54a}$$

and

$$Ax_0 - bN^*(cx_0) = 0. \tag{54b}$$

Remark. Theorem 4.1 shows that there exists a large class of systems with great practical importance for which the crucial assumption given by (9) of Theorem 3.1 is satisfied.

Proof of Theorem 4.1. Suppose (9) is not satisfied. Then by compactness of the interval $[0, T]$ the set $\mathscr{H}$ given by (8) has a limit point $\bar{t}$ in $[0, T]$. It follows from continuity of $c\dot{x}^*(\cdot)$ that $\mathscr{H}$ is a closed set, so we necessarily have $\bar{t}$ in $\mathscr{H}$; or equivalently

$$c\dot{x}^*(\bar{t}) = 0. \tag{55}$$

Let

$$\bar{x} \triangleq x^*(\bar{t}). \tag{56}$$

We construct a solution for (53a) in a neighbourhood of $\bar{t}$ with an initial condition $\bar{x}$ using Picard's successive approximations. By uniqueness this solution coincides with $x^*(t)$ in this neighbourhood, say, $(\bar{t} - \xi, \bar{t} + \xi)$. The Picard iteration is given by

$$x_0(t) = \bar{x}, \tag{57a}$$

$$x_p(t) = \bar{x} + \int_{\bar{t}}^{t} [Ax_{p-1}(\tau) + bN(cx_{p-1}(\tau))]\, d\tau, \tag{57b}$$

$$p = 1, 2, \ldots, \forall t \in (\bar{t} - \xi, \bar{t} + \xi).$$

The error between the pth iteration and the actual solution is given as follows ([1], p. 13). $\exists$ a constant M such that for all positive integers p and all $t \in (\bar{t} - \xi, \bar{t} + \xi)$,

$$| x^*(t) - x_p(t)| \leqslant \frac{M(| t - \bar{t} |)^{p+1}}{(p+1)!}. \tag{58}$$

Differentiating (57b) with respect to t and using Lipschitz properties of N it can be shown that $\exists$ a constant M' such that for all positive integers p and all $t \in (\bar{t} - \xi, \bar{t} + \xi)$

$$| c\dot{x}^*(t) - c\dot{x}_p(t)| \leqslant \frac{M'(| t - \bar{t} |)^p}{p!}. \tag{59}$$

Using (55) and (53a) we obtain

$$c[A\bar{x} + bN(c\bar{x})] = 0. \tag{60}$$

Using (57b) with $p = 1$,

$$x_1(t) = \bar{x} + \int_{\bar{t}}^{t} [A\bar{x} - bN(c\bar{x})]\, d\tau,$$

and

$$x_1(t) = \bar{x} + (t - \bar{t})\,[A\bar{x} - bN(c\bar{x})]. \tag{61}$$

Observe that using (60) in (61) we have that

$$cx_1(t) = c\bar{x}, \qquad \forall t \in (\bar{t} - \xi, \bar{t} + \xi). \tag{62}$$

Substituting (61) and (62) into (57b) with $p = 2$ we obtain the following equation

$$\dot{x}_2(t) = A\bar{x} + (t - \bar{t})\, A[A\bar{x} - bN(c\bar{x})] - bN(c\bar{x}). \tag{63}$$

Again using (60) in (63)

$$c\dot{x}_2(t) = (t - \bar{t})\, cA[A\bar{x} - bN(c\bar{x})]. \tag{64}$$

Using (64) the following relation is obtained

$$c\dot{x}^*(t) = (t - \bar{t})\, cA[A\bar{x} - bN(c\bar{x})] + (c\dot{x}^*(t) - c\dot{x}_2(t)). \tag{65}$$

Dividing (65) by $(t - \bar{t})$

$$\frac{c\dot{x}^*(t)}{(t - \bar{t})} = cA[A\bar{x} - bN(c\bar{x})] + \frac{(c\dot{x}^*(t) - c\dot{x}_2(t))}{(t - \bar{t})}. \tag{66}$$

Since in view of (59) the second term in the rightside of (66) can be made arbitrarily small for $| t - \bar{t} |$ small it follows that

$$cA[A\bar{x} - bN(c\bar{x})] = 0. \tag{67}$$

Otherwise we can find a neighbourhood of $\bar{t}$ such that $c\dot{x}^*(t) \neq 0$ for all t in this neighbourhood except at $t = \bar{t}$. This contradicts $\bar{t}$ being a limit point of $\mathscr{H}$.

It can easily be seen that proceeding in this way the following relation will be obtained

$$cA^i(A\bar{x} - bN(c\bar{x})) = 0, \qquad \forall i = 0, 1 \cdots n - 1. \tag{68}$$

But by the assumption of observability we have that the set of vectors c, cA, $\cdots$ cA^{n-1} span R^n [9 p. 502]. So we must have

$$A\bar{x} - bN(c\bar{x}) = 0, \tag{69}$$

or using (53a)

$$\dot{x}^*(\bar{t}) = 0, \tag{70}$$

which proves that $\bar{x}$ was an equilibrium point of (53a). Hence the theorem is proved.

5. Sector Conditions

In this section we restrict that set $\mathcal{N}$ by putting on it sector constraints. More specifically we define a new set of nonlinearities $\mathcal{N}_K$ as follows:

Definition. $N \in \mathcal{N}_K$ iff

(1) $N \in \mathcal{N}$;

(2) For a given constant $K > 0$,

$$-Kz \leqslant N(z) \leqslant Kz, \qquad \forall z \in [a, b]. \tag{71}$$

Note that for $K \geqslant S$ the sets $\mathcal{N}$ and $\mathcal{N}_K$ coincide. So in order to restrict $\mathcal{N}$ we let K to be smaller than S. It is assumed that $\mathcal{N}_K$ has the same topology as $\mathcal{N}$ given by (7).

We investigate the same problem as formulated in Section 2 except that $\mathcal{N}$ is replaced by $\mathcal{N}_K$. The necessary conditions of optimality for this problem is given by the following theorem.

Theorem 5.1. *Suppose N^* furnishes a local minimum for P on $\mathcal{N}_K$. Suppose that the assumption given by* (9) *of Theorem 3.1 is satisfied. Under these conditions N^* satisfies the following relations for all z (except possibly a finite set) in $[a, b]$.*

Case I $z \geqslant 0$.

(i) if $N^*(z) \neq \pm Kz$,

$$\frac{d^+N^*(z)}{dz} = -S \operatorname{sgn}\left[\int_{I(z, z_+(z))} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau\right]. \tag{72}$$

(ii) if $N^*(z) = +Kz$,

$$\frac{d^+N^*(z)}{dz} = -\left(\frac{K+S}{2}\right) \operatorname{sgn}\left[\int_{I(z, z_+(z))} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau\right] + \left(\frac{K-S}{2}\right), \tag{73}$$

where

$$I(z, v) \triangleq \{t \in [0, T];\ cx^*(t) \in (z, v)\}. \tag{74}$$

(iii) if $N^*(z) = -Kz$,

$$\frac{d^+N^*(z)}{dz} = -\left(\frac{K+S}{2}\right) \operatorname{sgn}\left[\int_{I(z, z_+(z))} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau\right] - \left(\frac{K-S}{2}\right), \tag{75}$$

where

$$\begin{aligned} z_+(z) \triangleq \operatorname{Min}\{y \in [z, b];\ & N^*(y) = Ky \text{ or } N^*(y) = -Ky, \\ & \text{and } \exists y' \in (z, y) \ni N^*(y') \neq \pm Ky'\}. \end{aligned} \tag{76a}$$

In case the set on the rightside of (76a) *is empty we take*

$$z_+(z) = b. \tag{76b}$$

To help visualize $z_+(z)$, note that $d^+N^*(z)/dz < K$, then the graph of N^* restricted to $(z, z_+(z))$ is interior to the sector.

Case II $z \leqslant 0$.

(i) if $N^*(z) \neq \pm Kz$,

$$\frac{d^-N^*(z)}{dz} = S \operatorname{sgn}\left[\int_{I(z_-(z), z)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau\right]. \tag{77}$$

(ii) if $N^*(z) = +Kz$,

$$\frac{d^-N^*(z)}{dz} = \left(\frac{K+S}{2}\right) \operatorname{sgn}\left[\int_{I(z_-(z), z)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right] d\tau\right] + \left(\frac{K-S}{2}\right). \tag{78}$$

(iii) if $N^*(z) = -Kz$,

$$\frac{d^-N^*(z)}{dz} = \left(\frac{K+S}{2}\right) \operatorname{sgn} \int_{I(z_-(z), z)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau - \left(\frac{K-S}{2}\right), \tag{79}$$

where

$$\begin{aligned} z_-(z) = \operatorname{Max}\{y \in [a, z];\ & N^*(y) = +Ky \text{ or } N^*(y) = -Ky, \\ & \text{and } \exists y' \in (y, z) \ni N^*(y') = \pm Ky'\}. \end{aligned} \tag{80a}$$

If again the set on the rightside of (80a) *is empty we take*

$$z_-(z) = a. \tag{80b}$$

All the other symbols are as given in Theorem 3.1.

Remark. Theorem 5.1 is a generalization of Theorem 3.1. It asserts that unless the argument of sgn function is zero, the slope of the optimum nonlinearity is $+S$ or $-S$ at the points interior to the sector; on the boundaries of the sector, $+S$ is replaced by $+K$ or $-S$ is replaced by $-K$ so that sector conditions are not violated.

Proof of Theorem 5.1. The theorem will only be proved for case I(ii). For the other cases the proof only requires simple changes. Consider the case where $z_+(z) < b$[4] and

$$\int_{I(z,z_+(z))} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau \neq 0, \tag{81}$$

then by continuity of $cx^*(\cdot)$ and assumption given by (9) there exists an open interval $(z_+(z) - \rho, z_+(z) + \rho)$ such that

$$\operatorname{sgn}\left[\int_{I(z,\alpha)}^{T} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt\right] = \operatorname{sgn}\left[\int_{I(z,z_+(z))}^{T} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt\right], \quad \forall \alpha \in (z_+(z) - \rho, z_+(z) + \rho), \tag{82}$$

(76a) implies that

$$N^*(z_+(z)) = + Kz_+(z), \tag{83}$$

or

$$N^*(z_+(z)) = - Kz_+(z). \tag{84}$$

First assume that (83) holds then there exists a z_1 in $(z_+(z) - \rho, z_+(z))$ (see (82)) such that

(a) $N^{*\prime}(z_1) > -S$ and $N^{*\prime}$ is continuous at z_1,

(b) $cx^*(t) \neq z_1$, $\quad \forall t \in \mathscr{H}$, $\mathscr{H}$ is given by (8).

Also there exists a z_2 in $(z_+(z), z_+(z) + \rho)$ such that

(a′) $N^{*\prime}(z_2) < +S$ and $N^{*\prime}$ is continuous at z_2,

(b′) $cx^*(t) \neq z_2$, $\quad \forall t \in \mathscr{H}$.

If (84) holds then $>$ and $-S$ are replaced by $<$ and $+S$, respectively, in (a) and vice-versa in (a′). We assume from here on that (83) holds. The modifications for the other case is straightforward.

[4] If $z_+(z) = b$ the proof reduces to that of Theorem 3.1 by taking k of $\tilde{N}_{k,\epsilon,z}$ between $+K$ and $-S$.

Let L_1 and L_2 be straight lines in $R \times R$ defined by the following equations

$$L_1 : \theta_1(v) = N^*(z_1) + B_1(v - z_1), \qquad \forall v \in R, \tag{85a}$$

and

$$L_2 : \theta_2(v) = N^*(z_2) + B_2(v - z_2), \qquad \forall v \in R, \tag{85b}$$

where B_1 and B_2 are any constants that satisfy the relations

$$-S < B_1 < N^{*\prime}(z_1) \tag{86a}$$

and

$$N^{*\prime}(z_2) < B_2 < +S. \tag{86b}$$

The next result is a direct consequence of the Implicit Function Theorem. There exists numbers $\gamma_1 > 0$ and $\gamma_2 < 0$ such that for each ξ in $(0, \gamma_1)$ there is a unique number $g(\xi)$ in $(\gamma_2, 0)$ such that

$$N^*(z_1 + g(\xi)) - N^*(z_1) = B_2 g(\xi) - \xi. \tag{87}$$

Similarly there exists numbers $\gamma_1' < 0$ and $\gamma_2' < 0$ such that for each ξ' in $(\gamma_1', 0)$ there is a unique number $\bar{g}(\xi')$ in $(\gamma_2', 0)$ such that

$$N^*(z_2 + \bar{g}(\xi')) - N^*(z_2) = B_2 \bar{g}(\xi') - \xi'. \tag{88}$$

To apply the Implicit Function Theorem for obtaining (87) we define the following function

$$F(\xi, \psi) \triangleq N^*(z_1 + \psi) - N^*(z_1) - B_1\psi + \xi. \tag{89}$$

Observe that F maps $R \times R$ into R and is continuously differentiable at $(0, 0)$ satisfying the following relations:

$$F(0, 0) = 0, \tag{90}$$

and

$$D_2F(0, 0) = N^{*\prime}(z_1) - B_1 .^5 \tag{91}$$

Using (86a)

$$D_2F(0, 0) > 0. \tag{92}$$

So hypothesis of the Implicit Function Theorem is satisfied by above relations and (87) follows. The reason for $g(\xi)$ to be negative (for $\xi > 0$) is that the Implicit Function Theorem requires

$$\frac{dg(\xi)}{d\xi}\bigg|_{\xi=0} = -(N^*(z_1) - B_1)^{-1} < 0. \tag{93}$$

The Eq. (88) is proved similarly.

[5] D_2 stands for partial derivative of F with respect to its second argument.

We now construct a perturbed nonlinearity $\tilde{N}_{\epsilon,k,z}$ as follows (see Fig. 4). If $k \in (d^+N^*(z)/dz, +K]$, then

$$\tilde{N}_{k,\epsilon,z}(y) \triangleq \begin{cases} N^*(y), & a \leqslant y \leqslant z; \\ N^*(z) + k(y - z), & z < y \leqslant z + \epsilon; \\ N^*(y) + k\epsilon - (N^*(z + \epsilon) - N^*(z)), & \\ \qquad z + \epsilon < y \leqslant z_1 + g[k\epsilon - (N^*(z + \epsilon) - N^*(z))]; & \\ N^*(z_1) + B_1(y - z_1), & \\ \qquad z_1 + g[k\epsilon - (N^*(z + \epsilon) - N^*(z))] < y \leqslant z_1; & \\ N^*(y), & z_1 < y \leqslant b. \end{cases}$$

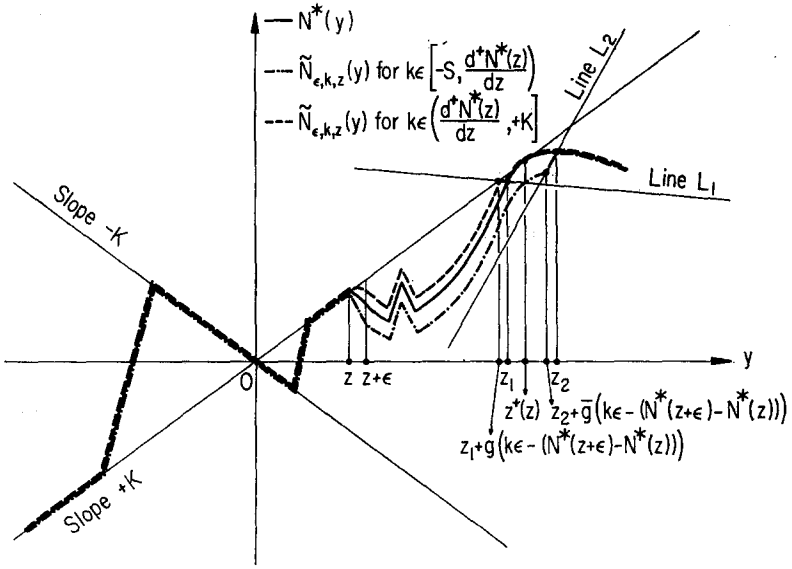

FIG. 4. The perturbed nonlinearity $\tilde{N}_{\epsilon,k,z}$ used in the proof of Theorem 5.1.

If $k \in [-S, (d^+N^*/dz))$, then

$$\tilde{N}_{k,\epsilon,z}(y) \triangleq \begin{cases} N^*(y), & a \leqslant y \leqslant z; \\ N^*(z) + k(y - z), & z < y \leqslant z + \epsilon; \\ N^*(y) + k\epsilon - (N^*(z + \epsilon) - N^*(z)), & \\ \qquad z + \epsilon < y \leqslant z_2 + \bar{g}[k\epsilon - (N^*(z + \epsilon) - N^*(z))]; & \\ N^*(z_2) + B_2(y - z_2), & \\ \qquad z_2 + \bar{g}[k\epsilon - (N^*(z + \epsilon) - N^*(z))] < y \leqslant z_2; & \\ N^*(y), & z_2 < y \leqslant b. \end{cases}$$

where ϵ is small enough to guarantee that

(i) $|k\epsilon - (N^*(z+\epsilon) - N^*(z))| < \min(\gamma_1, |\gamma_1'|)$,

(ii) $k\epsilon - (N^*(z+\epsilon) - N^*(z)) > 0 \qquad \text{if } k > \dfrac{d^+N^*(z)}{dz}$,

$k\epsilon - (N^*(z+\epsilon) - N^*(z)) < 0 \qquad \text{if } k < \dfrac{d^+N^*(z)}{dz}$,

(iii) $-Ky \leqslant \tilde{N}_{k,\epsilon,z}(y) \leqslant Ky, \qquad \forall y \in [a, b]$.

The justification of existence of such an upper bound for ϵ is straightforward, so that if the conditions stated above are satisfied

$$\tilde{N}_{k,\epsilon,z} \in \mathcal{N}_K. \tag{94}$$

We remark here that the perturbation of N^* constructed above is similar to that of Theorem 3.1 except that the perturbed nonlinearity meets N^* at z_1 (or z_2, depending on k) and follows N^* for $y > z_1$, where meeting N^* at z_1 is done through the straight line L_1 (or L_2, depending again on k). The point z_1, z_2 and the straight lines L_1 and L_2 are chosen such that $\tilde{N}_{\epsilon,k,z}$ remains in $\mathcal{N}_K$ for small ϵ and the contribution of the portions of $\tilde{N}_{\epsilon,k,z}$ on $(z, z+\epsilon)$ and on the line L_1 to the perturbed trajectory are of second order in ϵ.

Using the same procedure as in the proof of Theorem 3.1, that is, making a similar construction around the point z_1 (or z_2 depending on value of k) as well as z the following relations can be obtained:

$$P(\tilde{N}_{\epsilon,k,z}) - P(N^*) = \epsilon\left(k - \frac{d^+N^*(z)}{dz}\right)\int_{I(z,z_1)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau + o(\epsilon),$$

$$\forall k \in \left(\frac{d^+N^*(z)}{dz}, +K\right); \tag{95}$$

$$P(\tilde{N}_{\epsilon,k,z}) - P(N^*) = \epsilon\left(k - \frac{d^+N^*(z)}{dz}\right)\int_{I(z,z_2)} (\lambda^*(\tau))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} d\tau + o(\epsilon),$$

$$\forall k \in \left[-S, \frac{d^+N^*(z)}{dz}\right); \tag{96}$$

where we have used the fact that

$$\lim_{\epsilon\to 0} \frac{g[k - (N^*(z+\epsilon) - N^*(z))]}{\epsilon} < \infty \tag{97}$$

and

$$\lim_{\epsilon\to 0} \frac{\bar{g}[k\epsilon - (N^*(z+\epsilon) - N^*(z))]}{\epsilon} < \infty, \tag{98}$$

which follows from (93) and the corresponding relation for $\bar{g}$.

The inequalities (97) and (98) are used to show that

$$\overline{\lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}}} \frac{| N_{k,\epsilon,z} - N^* |_\infty}{\epsilon} < \infty, \tag{99}$$

and to show that the remainder terms are of $o(\epsilon)$ type.

Using (82) and the choice of z_1 and z_2, (95) and (96) reduces to the following relation for small enough ϵ so that $o(\epsilon)$ terms may be omitted:

$$\begin{aligned} &\operatorname{sgn}(P(\tilde{N}_{\epsilon,k,z}) - P(N^*)) \\ &\quad = \operatorname{sgn}\left[\left(k - \frac{d^+N^*(z)}{dz}\right) \int_{I(z,z_+(z))} (\lambda^*(t))^T \left[\frac{\partial f}{\partial N}\right]_{\text{opt}} dt\right], \\ &\qquad \forall k \in [-S, +K], \qquad k \neq \frac{d^+N^*(z)}{dz}, \end{aligned} \tag{100}$$

which implies that we have

$$P(\tilde{N}_{\epsilon,k,z}) < P(N^*),$$

unless (73) is satisfied, which contradicts optimality of N^*.

This completes proof of Theorem 5.1.

6. Extensions

In this section we point out some simple extensions of the theory developed previously and we state, without proof, a theorem for discrete dynamical systems.

I. If the right hand side of the nonlinear differential Eq. (1) is of the form $f(x, N(cx), t)$, then we add $x' \triangleq t$ as a new state variable and consider the augmented system

$$\begin{bmatrix} \dot{x} \\ \dot{x}' \end{bmatrix} = \begin{bmatrix} f(x, N(cx), x') \\ 1 \end{bmatrix}. \tag{101}$$

If for each x and N, $f(x, N(cx), \cdot)$ is continuously differentiable on $[0, T]$ and f is Lipschitz in x and $N(\cdot)$ (uniformly in $t \in [0, T]$) it can be shown that the rightside of (101) satisfies the conditions stated in Section 2, so that the results obtained previously are also valid for such time-varying dynamical systems.

II. If the performance index is of the form[6]

$$P(N) = \int_0^T h(x(t), N(cx(t)))\, dt, \tag{102}$$

then, all the previous results are valid if we replace the integrand $(\lambda^*(t))^T [\partial f/\partial N]_{\text{opt}}$ by $(\lambda^*(t))^T [\partial f/\partial N]_{\text{opt}} + [\partial h/\partial N]_{\text{opt}}$, where it is assumed that h is continuously differentiable with respect to its second argument as well as first.

III. A careful investigation of proof of Theorem 5.1 suggests that constraints more general than sector constraints can be considered for which a similar form of Theorem 5.1 will hold. One such example of practical value is the saturation constraint. Namely we define $\mathcal{N}_L$ as $N(\cdot) \in \mathcal{N}_L$ iff

(i) $N \in \mathcal{N}$,

and

(ii) $|N(y)| \leqslant L \qquad \forall y \in R$,

where L is a given constant.

The result analogous to Theorem 5.1, roughly asserts that the optimal nonlinearity uses either full slope or follows the saturation line with zero derivative, depending on the argument of a sgn function.

IV. Consider a discrete dynamical system described by the following nonlinear difference equation

$$x_{i+1} - x_i = f(x_i, N(cx_i)), \qquad i = 0, 1, ..., I, \tag{103}$$

where $f(\cdot, \cdot)$ is assumed to be continuously differentiable on $R^k \times R$ and x_0 is a fixed vector in R^n.

The constraint set $\mathcal{N}_D$ for the nonlinearities is defined as follows $N \in \mathcal{N}_D$ iff

(i) $$N \in C^{(1)}, \tag{104}$$

and

(ii) $$-Ky \leqslant N(y) \leqslant +Ky, \qquad \forall y \in R. \tag{105}$$

[6] Such forms may occur as a result of introducing penalty functions to insure that $N^*(\cdot)$ is near enough to a given nonlinearity.

We define a performance criterion as

$$P(N) = \sum_{i=0}^{I} h(x_i^N), \tag{106}$$

where x_i^N is the trajectory of (103) corresponding to $N \in \mathcal{N}_D$ and $h \in C^{(1)}$.

The minimization problem can be stated as

$$\underset{N \in \mathcal{N}_D}{\text{Min}}\, P(N). \tag{107}$$

We assume that the topology on $\mathcal{N}_D$ is the one induced by the sup norm as given by (7). The following theorem gives the necessary conditions of optimality.

THEOREM 6.1. *Suppose $N^* \in \mathcal{N}_D$ furnishes a local minimum for P; then the following relations hold:*

$$\begin{aligned} N^*(\alpha_m) &= -K \operatorname{sgn}\left(\sum_{j \in J(\alpha_m)} (\lambda_j^*)^T \frac{\partial f}{\partial N_{\text{opt}}}(j) \right) \alpha_m , \\ &\forall m = 1,..., M, \end{aligned} \tag{108}$$

where

$$\{\alpha_m\}_{m=1}^{M} \triangleq (y \in R;\ \exists j \in (0, 1,..., I),\ cx_j^* = y), \tag{109}$$

and

$$J(\alpha_m) \triangleq (j \in (0, 1,..., I);\ cx_j^* = \alpha_m). \tag{110}$$

$\{\lambda_i^*\}_{i=0}^{I}$ *solves the adjoint equation given by*

$$\lambda_i^* - \lambda_{i+1}^* = \left[\frac{df(x_i^*, N^*(cx_i^*))}{dx_i^*}\right]^T \lambda_{i+1}^* + \left[\frac{dh(x_i^*)}{dx_i^*}\right]^T, \tag{111}$$

with $\lambda_I^* = 0$.

APPENDIX

In order to prove Lemma 3.1 to Lemma 3.3 we need the following auxiliary Lemmas.

LEMMA A1. (a) $\lim_{\epsilon \to 0} \| \tilde{x} - x^* \| = 0$;

(b) $\lim_{\epsilon \to 0} \| \dot{\tilde{x}} - \dot{x}^* \| = 0$.

LEMMA A2. (a) $\dot{y}^*(\omega_{2i-1}) > 0$, $i = 1,..., \ell$;

(b) $\dot{y}^*(\omega_{2i}) < 0$, $i = 1,..., \ell$.

LEMMA A3. *Consider the nonlinear differential equation*

$$\dot{w} = g(w, t), \tag{A1}$$

where g is Lipschitz in w (uniformly in t) and for any continuous function $w(t)$, $g(w(t), t)$ is locally integrable in t.

Let $w_1(t)$ and $w_2(t)$ be approximate solutions to (A1) *on the interval $[t_1, t_2]$ in the sense that*

$$|\dot{w}_1(t) - g(w_1(t), t)| < \epsilon_1(t), \tag{A2a}$$

and

$$|\dot{w}_2(t) - g(w_2(t), t)| < \epsilon_2(t), \tag{A2b}$$

for almost all t in $[t_1, t_2]$ with $w_1(t_1) = w_2(t_1)$. Then

$$|w_1(t) - w_2(t)| < \left[\int_{t_1}^{t_2} (\epsilon_1(\tau) + \epsilon_2(\tau))\, d\tau\right] \exp L(t_2 - t_1), \qquad \forall t \in [t_1, t_2], \tag{A3}$$

where L is the Lipschitz constant of g.

The proofs for Lemma A2 and Lemma A3 are left to the reader. Lemma A3 is a modified version of 10.5.1 p. 282 [2].

Proof of Lemma A1. Substituting $\tilde{x}(t)$ and $x^*(t)$ into (1)

$$\dot{\tilde{x}}(t) - \dot{x}^*(t) = f(\tilde{x}(t), \tilde{N}(c\tilde{x}(t))) - f(x^*(t), N^*(cx^*(t))). \tag{A4}$$

Integrating both sides of (A4), taking norms and using Lipschitzness of f we obtain

$$\begin{aligned} |\tilde{x}(t) - x^*(t)| \leqslant &\int_0^t K_1 |\tilde{x}(\tau) - x^*(\tau)|\, d\tau \\ &+ \int_0^t K_2 |(\tilde{N}(c\tilde{x}(\tau)) - N^*(cx^*(\tau))|\, d\tau. \end{aligned} \tag{A5}$$

Adding and subtracting $\tilde{N}(cx^*(\tau))$ inside the absolute value in the second integral and then using properties of $\mathcal{N}$ we have

$$\begin{aligned} |\tilde{x}(t) - x^*(t)| \leqslant &\int_0^t K_1 |\tilde{x}(\tau) - x^*(\tau)|\, d\tau + \int_0^t K_2 S |c| |(\tilde{x}(\tau) - x^*(\tau)|\, d\tau \\ &+ \int_0^t K_2 |\tilde{N}(cx^*(\tau)) - N^*(cx^*(\tau))|\, d\tau, \end{aligned} \tag{A6}$$

$$|\tilde{x}(t) - x^*(t)| \leqslant \int_0^t (K_1 + K_2 S |c|) |\tilde{x}(\tau) - x^*(\tau)|\, d\tau + K_2 T |\tilde{N} - N^*|_\infty. \tag{A7}$$

Applying the Bellman-Gronwall inequality

$$\| \tilde{x} - x^* \| \leqslant K_2 T \, | \tilde{N} - N^* |_\infty \exp(K_1 + K_2 S \, | \, c \, |) \, T. \tag{A8}$$

By using (24) part (a) follows. To prove part (b) we substitute (A8) into (A4) and use assumptions on f and $\mathcal{N}$ to obtain

$$\| \dot{x}^* - \dot{\tilde{x}} \| \leqslant K_2 \, | N^* - \tilde{N} |_\infty \, [T(K_1 + K_2 S \, | \, c \, |) \exp(K_1 + K_2 S \, | \, c \, |) \, T + 1]. \tag{A9}$$

Proof of Lemma 3.1. We first make the following definition

$$\eta \triangleq \min_{j=1,\ldots m} | \, z - cx^*(t_j) | \, , \tag{A10}$$

where $t_j \in \mathcal{H}$, $j = 1,\ldots, m$, and $\mathcal{H}$ is as defined by (8). It follows by assumption given by (17) that η is a positive number, so the hypothesis of the Inverse Function Theorem is satisfied by continuous differentiability of $y^*(\cdot)$. So for each ω_i there exists an interval I_i containing z and a continuously differentiable function $T_i(\cdot)$ such that

$$T_i : I_i \to N_i \, , \qquad T_i(y^*(\omega)) = \omega, \qquad \forall \omega \in N_i \, , \tag{A11}$$

where N_i is an interval containing ω_i such that

$$y^*(N_i) = I_i \, , \qquad N_i \cap N_j = \varnothing \, , \qquad \forall i, j = 1,\ldots, 2\ell, \tag{A12}$$

$$\operatorname{sgn} \dot{y}^*(\omega) = \operatorname{sgn} \dot{y}^*(\omega_i), \qquad \forall \omega \in N_i \, . \tag{A13}$$

We pick the number β satisfying the following conditions

$$0 < \beta < \frac{\eta}{2} \, , \tag{A14}$$

$$[z - \beta, z + 2\beta] \subset \bigcap_{i=1}^{2\ell} I_i \, , \tag{A15}$$

where η is given by (A10).

By Lemma A1 for $\beta/2 > 0$, there exists an ϵ'' such that

$$\| \tilde{y}(\cdot) - y^*(\cdot) \| < \frac{\beta}{2} \, , \qquad \forall \epsilon < \epsilon''. \tag{A16}$$

Choose $\epsilon'(\beta)$ such that

$$\epsilon'(\beta) = \min(\epsilon'', \beta). \tag{A17}$$

Then by (A14) and (A15) we obtain the following relations

$$\beta + \epsilon'(\beta) < \eta, \tag{A18}$$

$$[z - \beta, z + \epsilon'(\beta) + \beta] \subset \bigcap_{i=1}^{\ell} I_i . \tag{A19}$$

We now make the following definitions for each $i = 1,..., 2\ell$:

$$s_i^+ \triangleq T_i(z + \epsilon'(\beta) + \beta), \tag{A20}$$

and

$$s_i^- \triangleq T_i(z - \beta). \tag{A21}$$

It is easy to show by using (A12) and (A18) that

$$y^*(s) = z + \epsilon'(\beta) + \beta \quad \text{iff} \quad s = s_j^+ \quad \text{for some} \quad j \in (1,..., 2\ell), \tag{A22a}$$

and

$$y^*(s) = z - \beta \quad \text{iff} \quad s = s_j^- \quad \text{for some} \quad j \in (1,..., 2\ell). \tag{A22b}$$

The continuity of each T_i proves (25a). The order on s_i's given by (27) follows from Lemma A2, (A13) and the following property of each T_i.

$$\frac{dT_i(v)}{dv} = [\dot{y}^*(T_i(v))]^{-1}, \qquad \forall v \in I_i . \tag{A23}$$

In order to prove part (b), by using (A16) and (A17) we observe that

$$\tilde{y}(t) - \beta/2 < y^*(t) < \tilde{y}(t) + \beta/2, \qquad \forall t \in [0, T], \qquad \forall \epsilon \leqslant \epsilon'(\beta). \tag{24}$$

By inequality (A24) it is enough to prove the following relations.

$$y^*(t) > z + \epsilon'(\beta) + \beta, \qquad \forall t \in \bigcup_{i=1}^{\ell} (s_{2i-1}^+ , s_{2i}^+), \tag{A25a}$$

and

$$y^*(t) < z - \beta, \qquad \forall t \in [0, s_1^-) \cup (s_{2\ell}^- , T] \cup \left\{ \bigcup_{i=1}^{\ell-1} (s_{2i}^- , s_{2i+1}^-) \right\}. \tag{A25b}$$

This is proved by contradiction as follows: Suppose there exists t in (s_{2i-1}^+ , s_{2i}^+) such that

$$y^*(t) \leqslant z + \epsilon'(\beta) + \beta. \tag{A26}$$

Using Lemma A2 and (A13) we have that

$$\dot{y}(s_{2i-1}^+) > 0. \tag{A27}$$

So

$$\exists t' : s_{2i-1}^{+} < t' < t < s_{2i}^{+}, \qquad \text{and} \qquad y^*(t') > z + \epsilon'(\beta) + \beta.$$

By the Intermediate Value Theorem there exist a t'' such that

$$t'' \in (t', s] \tag{A28}$$

and

$$y(t'') = z + \epsilon'(\beta) + \beta. \tag{A29}$$

(A29) contradicts (A22a), so (A26) cannot be true. This proves (A25a). The proof of (A25b) is similar.

Proof of Lemma 3.2. It is enough to show that

$$\int_{s_{2i-1}^{-}}^{s_{2i-1}^{+}} (\dot{x}^*(\tau) - \dot{\hat{x}}(\tau))\, d\tau = o(\epsilon, \beta) \tag{A30}$$

and

$$\int_{s_{2i}^{+}}^{s_{2i}^{-}} (\dot{x}^*(\tau) - \dot{\hat{x}}(\tau))\, d\tau = o(\epsilon, \beta). \tag{A31}$$

Using (A9), (24) and (26a) the result follows.

Proof of Lemma 3.3. If $N^*(\cdot)$ were continuously differentiable the result would be a well known property of differential equations. By definition of $\mathcal{N}$, N^* has a derivative with finite number of discontinuities on $[a, b]$ denoted by $\{\alpha_i\}_{i=1}^{K}$.

We define the inverse image of α_i for each i as follows

$$\Lambda_i \triangleq (t \in [0, T]; y^*(t) = \alpha_i). \tag{A32}$$

By (9) Λ_i is a finite set for each i, so we have

$$\mu\left(\bigcup_{i=1}^{K} \Lambda_i\right) = 0, \tag{A33}$$

where μ is the Lebesgue measure on real line.

We now define for each i a sequence of decreasing sets as follows

$$\Lambda_i^j \triangleq \left(t; y^*(t) \in \left(\alpha_i - \frac{1}{j}, \alpha_i + \frac{1}{j}\right)\right), \qquad j = 1, 2, \ldots. \tag{A34}$$

By a well known result in measure theory [6, p. 61]

$$\lim_{j\to\infty} \mu\left(\bigcup_{i=1}^{K} \Lambda_i^j\right) = \mu\left(\bigcap_{j=1}^{\infty} \bigcup_{i=1}^{K} \Lambda_i^j\right) = 0. \tag{A35}$$

Define $\delta x(t)$ as

$$\delta x(t) \triangleq \bar{x}(t) - x^*(t). \tag{A36}$$

Using Eq. (1) we obtain the following relation

$$\dot{\delta x}(t) = f(x^*(t) + \delta x(t), N^*(c(x^*(t) + \delta x(t))) + \gamma) - f(x^*(t), N^*(cx^*(t))),$$
$$\forall t \in [t', T], \qquad \delta x(t') = \delta x'. \tag{A37}$$

By using the Lipschitz conditions on f and N^* it can be shown that $\exists$ constants C_1 and C_2 such that:

$$\| \delta x \| < C_1(| \delta x' | + | \gamma |)^7 \tag{A38a}$$

and

$$\| \dot{\delta x} \| < C_2(| \delta x' | + | y |). \tag{A38b}$$

By (A38a) for each $1/2j > 0$, j a positive integer, there is a number $\delta(j) > 0$:

$$\| c\delta x \| < \frac{1}{2j}, \qquad \forall(| \delta x' | + | \gamma |) < \delta(j). \tag{A39}$$

If $t \notin \bigcup_{i=1}^{K} \Lambda_i^j$ then using definition given by (A34)

$$y^*(t) \notin \bigcup_{i=1}^{K} \left(\alpha_i - \frac{1}{j}, \alpha_i + \frac{1}{j}\right), \tag{A40}$$

so that using (A39) we have

$$y^*(t) + c\delta x(t) \notin \bigcup_{i=1}^{K} \left(\alpha_i - \frac{1}{2j}, \alpha_i + \frac{1}{2j}\right),$$
$$\forall(| \delta x' | + | \gamma |) < \delta(j), \qquad \forall t \in [t', T] \sim \bigcup_{i=1}^{K} \Lambda_i^j.^8 \tag{A41}$$

By continuity and compactness $dN^*(y)/dy$ is uniformly continuous on the set

$$[a, b] \sim \bigcup_{i=1}^{K} \left(\alpha_i - \frac{1}{j}, \alpha_i + \frac{1}{j}\right).$$

[7] Here $\| \cdot \|$ is computed by taking supremum over the interval $[t', T]$.

[8] $\sim$ denotes difference of two sets.

So that we may expand (A37) around the optimal trajectory for $t \notin \bigcup_{i=1}^{K} \Lambda_i^j$ as follows

$$\delta\dot{x}(t) = \left[\frac{df(x, N^*(cx))}{dx}\right]_{x=x^*(t)} \delta x(t) + \left[\frac{\partial f(x, u)}{\partial u}\right]_{(x,u)=(x^*(t,N^*(cx^*(t)))} \gamma + o(|\delta x(t)| + |\gamma|, t), \tag{A42}$$

$$\forall t \in [t', T] \sim \bigcup_{i=1}^{K} \Lambda_i^j, \qquad (|\delta x'| + |\gamma|) < \delta(j).$$

Using (A38a) we can write (A42) as follows:

$$\delta\dot{x}(t) = \frac{df(x, N^*(cx))}{dx}\Big|_{x=x^*(t)} \delta x(t) + \frac{\partial f(x, u)}{\partial u}\Big|_{(x,u)=(x^*(t),N^*(cx^*(t)))} \gamma + o(|\delta x'| + |\gamma|, t), \tag{A43}$$

$$\forall t \in [t, T] \sim \bigcup_{i=1}^{K} \Lambda_i^j, \qquad (|\delta x'| + |\gamma|) < \delta(j),$$

where uniform continuity of $dN^*(y)/dy$ insures that $o(|\delta x'| + |\gamma|, t)$ is uniform in t.

We now consider the linear differential equation given below

$$\dot{v}(t) = \left[\frac{df(x, N^*(cx))}{dx}\right]_{x=x^*(t)} v(t) + \left[\frac{\partial f(x, u)}{\partial u}\right]_{(x,u)=(x^*(t),N^*(cx^*(t)))} \gamma, \tag{A44}$$

where

$$v(t') = \delta x', \qquad t \in [t', T].$$

It can easily be shown that $\exists$ constants C_1' and C_2':

$$\|v\| < C_1'(|\delta x'| + |\gamma|) \tag{A45a}$$

and

$$\|\dot{v}\| < C_2'(|\delta x'| + |\gamma|). \tag{A45b}$$

Comparing Eqs. (31) and (33) with (A43) we observe that it is enough to prove the following relation

$$\lim_{|\delta x'|+|\gamma|\to 0} \frac{\|\delta x - v\|}{|\delta x'| + |\gamma|} = 0. \tag{A46}$$

Let $I_{\Gamma_1(j)}$ and $I_{\Gamma_2(j)}$ be the indicator functions of the sets $\Gamma_1(j)$ and $\Gamma_2(j)$ where

$$\Gamma_1(j) \triangleq \bigcup_{i=1}^{K} \Lambda_i^j \tag{A47}$$

and

$$\Gamma_2(j) \triangleq [t', T] \sim \Gamma_1(j). \tag{A48}$$

Now apply Lemma A3 with $g(w, t)$ defined as

$$\begin{aligned} g(w, t) = &\left[\frac{df(x, N^*(cx))}{dx}\right]_{x=x^*(t)} \cdot (I_{\Gamma_2(j)})\, w \\ &+ \left[\frac{\partial f(x, u)}{\partial u}\right]_{(x,u)=(x^*(t),N^*(cx^*(t)))} \cdot I_{\Gamma_2(j)} \cdot \gamma, \end{aligned} \tag{A49}$$

where we consider $\delta x(t)$ and $v(t)$ as the approximate solutions of (A49) on $[t', T]$. Consequently we have

$$\begin{aligned} \| \delta x - v \| < \Big(\int_{t'}^{T} &(| \delta \dot{x}(t)\, I_{\Gamma_2(j)} | + | o(| \delta x' | + | \gamma | , t)\, I_{\Gamma_2(j)} | \\ &+ | \dot{v}(t)\, I_{\Gamma_1(j)} |)\, dt \Big) \exp(M(T - t')), \end{aligned} \tag{A50}$$

where

$$M \triangleq \left\| \left[\frac{df(x, N^*(cx))}{dx}\right]_{x=x^*(t)} \right\|. \tag{A51}$$

Using uniformity of $o(| \delta x' | + | \gamma | , t)$ in t and relations (A38b) and (A45b), (A50) can be written as

$$\begin{aligned} \| \delta x - v \| < ((C_2 + C_2') \cdot (| \delta x' | + | \gamma |)\, \mu(\Gamma_1(j)) \\ + o(| \delta x' | + | \gamma |)) \exp M(T - t'); \end{aligned} \tag{A52}$$

(A46) then follows by using (A35) and (A47) in (A52) which proves the lemma.

References

1. E. A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations," McGraw-Hill, 1955.
2. J. Dieudonné, "Foundations of Modern Analysis," Academic Press, New York, 1960.

3. C. GOFFMAN AND G. PEDRICK, "First Course in Functional Analysis," Prentice Hall, 1965.
4. M. R. HESTENES, "Calculus of Variations and Optimal Control Theory," John Wiley & Sons, 1966.
5. E. B. LEE AND L. MARKUS, "Foundations of Optimal Control Theory," John Wiley & Sons, 1967.
6. H. L. ROYDEN, "Real Analysis," MacMillan, 1968.
7. K. L. SURYANARAYANAN AND A. C. SOUDACK, Analog computer automatic parameter optimization of nonlinear control systems with specific inputs, *IEEE Trans. on Computers*, August 1968.
8. J. WILLEMS, Nonlinear Harmonic Analysis, Ph.D. Thesis, Dept. of Electrical Engineering, Massachusetts Institute of Technology, 1968.
9. L. A. ZADEH AND C. A. DESOER, "Linear System Theory," McGraw-Hill, 1963.