ON THE PERTURBATIONAL SENSITIVITY OF SOLUTIONS TO NONLINEAR DIFFERENTIAL EQUATIONS

by

M. K. Inan

Memorandum No. ERL-M270

23 February 1970

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE PERTURBATIONAL SENSITIVITY OF SOLUTIONS TO NONLINEAR DIFFERENTIAL EQUATIONS*

M. K. Inan†

1. Introduction. An important problem that comes up very often in optimization theory is the perturbational behavior of the trajectory of an n-vector differential equation of the form

$$x(t) = f(x(t), u(t), t) \tag{1}$$

with respect to perturbations in the initial condition x_o and in the forcing (input) function $u(t)$. For a suitably chosen topology for the input function-space $\mathcal{U}$ (to which $u(\cdot)$ belongs), and the function-space $\mathcal{C}$ (to which trajectory $x(\cdot)$ belongs) the d.e. (1) induces a nonlinear operator mapping $\mathcal{U} \times R^n$ into $\mathcal{C}$ provided existence and uniqueness of solutions for those inputs $u(\cdot)$ and initial conditions x_o are guaranteed. The problem of interest is to give sufficient conditions (as weak as possible) under which the operator mentioned above is continuous and differentiable (in the sense of Fréchet) at a given point. Such sufficiency conditions have extensively been given in literature, some of which, for example, are references [1] to [5]. However the sufficiency conditions given in literature so far (to the author's knowledge) have been strict enough to exclude some important class of systems. For example systems containing piecewise-affine nonlinearities

---

*Received by

†Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, California, 94720. Research sponsored by the National Aeronautics and Space Administration Grant NGL-05-003-016(Sup 6) and the National Science Foundation Grant GK-10656X.

violate the generally assumed smoothness (differentiability) conditions.

To be more specific consider the following example of such a 'kinky' system. The d.e. of interest is given as

$$\dot{x}(t) = f(x(t), N(cx(t))) \tag{2}$$

where $f(.,.)$ is a C^1 mapping of $R^n \times R$ into R^n, c is a constant row vector and N maps R into R with a piecewise-continuous derivative. If α is a point of discontinuity of $\frac{dN(y)}{dy}$, then $\frac{\partial f}{\partial x}$ $(x, N(cx))$ is not defined on the hyperplane: $\{x \in R^n; cx = \alpha\}$. If there are a finite number of such hyperplanes does (2) possess a 'variational equation' describing the linearized perturbations of the trajectory corresponding to perturbations in the initial condition? or (in the operator-theoretic formulation as in the previous paragraph) is the initial-condition-to-trajectory map Fréchet differentiable? The answer to this question turns out to be affirmative, provided that the Lebesgue measure of the subset of the time interval at which the nominal trajectory remains on the discontinuity hyperplanes is zero [8]. Motivated by such examples of 'kinky' systems the sufficiency conditions stated in this paper are weakened to the extent that any further reasonable weakening would possibly result in the violation of necessary conditions of continuity and differentiability.

2. Formulation and Initial Assumptions. Let the input and trajectory spaces be defined respectively as

$$\mathcal{U} \triangleq L_\infty^m [0,T] \tag{3a}$$

$$\mathcal{C} \triangleq C_n [0,T] \tag{3b}$$

ASSUMPTION: Let f, as given in (1), map $R^n \times R^m \times R$ into R^n and assume that for a given pair (x_o, u^*) in $R^n \times \mathcal{U}$,(1) has a unique solution x^* in $\mathcal{C}$.

NOTATIONAL REMARKS. The symbol $|\cdot|$ will stand for the absolute value of a number, the norm of a vector in R^n (or R^m), or a compatible norm of a matrix. The vector and matrix norms used are given below

(4a) $$|x| \triangleq \sum_{i=1}^{n} |x_i| \quad , \quad x = \text{n-vector}$$

(4b) $$|A| \triangleq \sum_{\substack{i=1\ \ldots,n \\ j=1,\ldots,m}} |aij| \quad , \quad A = \text{mxn matrix}$$

The symbol $\|\cdot\|$ stands for $\operatorname{ess\,sup}_{[0,T]} |\cdot|$. $B^n(x,r)$ denotes a closed ball of radius r, centered at x with the norm given by (4a) in R^n. The norm on a product space is taken to be sum of the norms of component spaces.

Using the notation developed above, and the assumption, let $r > 0$ be a number such that

(5a) $$\|x^*\| < r$$

(5b) $$\|u^*\| < r$$

whenever the dependence of x^* on x_o and u^* is to be amplified x^* will be denoted by $\phi(\cdot, x_o, u^*)$.

3. <u>Continuity</u>.

THEOREM 1. Suppose that

C1] a) for each fixed (x,u) in $B^n(0,2r) \times B^m(0,2r)$, $f(x,u,\cdot)$ is measurable on [0,T],

(b) for each t in $[0,T]$, $f(\cdot,\cdot,t)$ is continuous on an open set containing $B^n(0,2r) \times B^m(0,2r)$,

C2] there exist non-negative integrable† real-valued functions $k_1(\cdot)$ and $k_2(\cdot)$ on $[0,T]$ such that for each t in $[0,T]$

$$|f(x_1 u_1,t) - f(x_2,u_2,t)| \leq k_1(t)\ |x_1-x_2| + k_2(t)\ |u_1-u_2|$$

$$\forall (x_1,u_1),\ (x_2,u_2) \in B^n(0,2r) \times B^m(0,2r),$$

and suppose that the existence of solution assumption of previous section is satisfied, then, there is a constant $\gamma > 0$ such that (1) has a unique solution $\phi(\cdot,x_o + \delta x_o,\ u^* + \delta u)$ in $\mathcal{C}$ corresponding to each pair $(x_o + \delta x_o,\ u^* + \delta u)$ satisfying the relation

(6a) $$|\delta x_o| + \|\delta u\| < \gamma$$

Moreover the solution in $\mathcal{C}$ is Lipschitz around (x_o,u^*) as given by the relation

(6b) $$\|\phi(\cdot,x_o,u^*) - \phi(\cdot,x_o + \delta x_o,\ u^* + \delta u)\| \leq Ke^{K_1}(|\delta x_o| + \|\delta u\|)$$

$$\forall (\ x_o,\delta u) \text{ satisfying (6a)}$$

where

$$K \triangleq \max(1,\int_0^T k_2(t)\ dt)$$

$$K_1 \triangleq \int_0^T k_2(t)\ dt$$

†A function is said to be integrable if it is measurable, and the integral of its absolute value (norm) is finite.

Before proceeding to the proof of Theorem 1 let us remark on a technicality for the purposes of unambiguity. Leave aside the proof, even a discussion of this technical point is neglected in most text books on differential equations due to its possibly elementary nature. The fact that has to be proved is the following:

For each $x \in C_n[0,T]$ and $u \in \mathcal{U}$

such that

$$\|x\| < 2r \quad , \quad \|u\| < 2r$$

$f(x(t), u(t), t)$ is integrable R^n valued function on [0,T]. We first have to show that each component of f is measurable in t. Taking $x_k(t)$ and $u_k(t)$ as sequences of simple functions converging almost everywhere to x(t) and u(t) we first prove measurability for fixed k([7] p. 85). Letting k go to infinity and using closedness of measurability under pointwise convergence ([6] p. 67), together with C1] b) which insures convergence, the measurability of $f(x(t), u(t),t)$ follows†. It remains to show that

$$\int_0^T |f(x(t), u(t),t)| \, dt < \infty$$

which follows from integrability of $f(x^*(t), u^*(t),t)$ and C2].

<u>Proof of Theorem 1.</u> Choose γ as follows

(7a) $$\gamma = \frac{re^{-K_1}}{2K}$$

Definitions of K and K_1 imply that

(7b) $$\gamma < r/2$$

---

†For more details of these type of proofs the reader may refer to any text-book on measure theory such as [6] or [7].

Let us for the time being assume for $(\delta x_o, \delta u)$ satisfying (6a) a unique solution exists on $[0,T]$ such that

(8) $$\|\phi(\cdot, x_o + \delta x_o, u^*(\cdot) + \delta u(\cdot))\| < 2r$$

For simplicity let

$$y(t) \triangleq \phi(t, x_o + \delta x_o, u^* + \delta u)$$

We then have

$$y(t) - x^* (t) = \delta x_o + \int_0^t [f(y(\tau), u^*(\tau) + \delta u(\tau),\tau) - f(x^*(\tau), u^*(\tau),\tau)]\, d\tau$$

Using (5), (7b), (8) and C2]

$$|y(t) - x^*(t)| \leq |\delta x_o| + \int_0^t k_1(\tau)|y(\tau) - x^*(\tau)|\, d\tau + \int_0^t k_2(\tau)|\delta u(\tau)|\, d\tau$$

$$|y(t) - x^*(t)| \leq |\delta x_o| + \|\delta u(\cdot)\| \int_0^t k_2(\tau)d\tau + \int_0^t k_1(\tau)|y(\tau) - x^*(\tau)|\, d\tau$$

Using Bellman-Gronwall Inequality

(9) $$|y(t) - x^*(t)| \leq Ke^{\int_0^t k_1(\tau)d\tau}(|\delta x_o| + \|\delta u\|) \leq Ke^{K_1}(|\delta x_o| + \|\delta u(\cdot)\|)$$

$$\forall\, t \in [0,T]$$

Using (6a) and (7a) in (9)

$$|y(t) - x^*(t)| \leq r/2 \quad , \quad \forall\, t \in [0,T]$$

or

(10) $$\|y(\cdot)\| \leq r/2 + \|x^*\| \leq 3r/2 < 2r$$

We now have to justify our initial assumption of existence of $\phi(t, x_o + \delta x_o, u^* + \delta u)$ in $B^n(0,2r)$ for $t \in (0,T)$. Let $[0,\xi]$ be the maximum interval† for which $\phi(t, x_o + \delta x_o, u^* + \delta u)$ is defined and is in $B^n(0,2r)$. It suffices to show that $\xi > T$. Suppose not, then there is a t' in $[0,T]$ such that

$$|\phi(t', x_o + \delta x_o, u^* + \delta u)| = 2r$$

and

$$|\phi(t, x_o + \delta x_o, u^* + \delta u)| \leq 2r, \forall t \in [0,t']$$

Using the same argument in deriving (10) we get

$$|\phi(t', x_o + \delta x_o, u^* + \delta u)| - |x^*(t')| \leq r/2.$$

or

$$|x^*(t')| \geq 2r - r/2 = 3r/2. \tag{11}$$

(11) clearly contradicts (5a) and the proof is complete

4. Differentiability. It has already been proved in the previous section that the map $(x, u(\cdot)) \longmapsto \phi(\cdot, x, u(\cdot))$ is well-defined in a neighborhood of (x_o, u^*), in particular, it is Lipschitz in this neighborhood. In this section we investigate the existence of the Fréchet derivative of the map mentioned above at the point (x_o, u^*) in $R^n \times \mathcal{U}$.

NOTATION

$$D_1 f(x,u,t) \triangleq \frac{\partial f}{\partial x}(x,u,t) \tag{12a}$$

---

†Existence and uniqueness of solutions to (1) on a nonempty interval follows from the local Lipschitz condition C2] and the condition C1].

(12b) $$D_2 f(x,u,t) \triangleq \frac{\partial f}{\partial u}(x,u,t)$$

For each t in [0,T], $D_1 f(x,u,t)$ and $D_2(x,u,t)$ are $n \times n$ and $n \times m$ matrices respectively.

ADDITIONAL ASSUMPTIONS: C3] $D_1 f(x^*(t), u^*(t), t)$ and $D_2 f(x^*(t), u^*(t), t)$ are well-defined for almost all t in [0,T] and

(13) $$\int_0^T (|D_1 f(x^*(t), u^*(t), t)| + |D_2 f(x^*(t), u^*(t), t)|)\, dt < \infty$$

C4] given $\varepsilon > 0$ and $\eta > 0$ there exist a set $\mathcal{S} \subset [0,T]$ and a number $\delta > 0$ such that

(a) $\mu(\mathcal{S}) < \eta$ ($\mu(\cdot)$ is the Lebesque measure on real line)

(b) $\forall\, t \in [0,T] \sim \mathcal{S}$ $D_1 f(x,u,t)$ and $D_2 f(x,u,t)$ are continuous on the ball centered at $(x^*(t), u^*(t)) \in R^n \times R^m$ with the radius $\delta > 0$.

(c) Given any $(\delta x, \delta u)$ in $L_\infty^n[0,T] \times \mathcal{U}$

such that

(14) $$\|\delta x\| + \|\delta u\| < \delta$$

The following hold

(15a) $$\int_{[0,T] \sim \mathcal{S}} |D_1 f(x^*(t), u^*(t), t) - D_1 f(x^*(t) + \delta x(t), u^*(t) + \delta u(t), t)|\, dt < \varepsilon$$

and

(15b) $$\int_{[0,T] \sim \mathcal{S}} |D_2 f(x^*(t), u^*(t), t) - D_2 f(x^*(t) + \delta x(t), u^*(t) + \delta u(t), t)|\, dt < \varepsilon$$

THEOREM 2. (Fréchet Derivative) Suppose that conditions C1] to C4] are satisfied, then the map

$$(x, u(\cdot)) \in R^n \times \mathcal{U} \longmapsto \phi(\cdot, x, u(\cdot)) \in \mathcal{C}$$

is Fréchet differentiable at (x_o, u^*). Its Fréchet derivative at $(x_o, u^*(\cdot))$ is the linear map

$$L : R^n \times \mathcal{U} \longrightarrow \mathcal{C}$$

given by

(16) $$L : (\delta x_o, \delta u) \longmapsto v(\cdot) = L(\delta x_o, \delta u)$$

where

(17a) $\dot{v}(t) = D_1 f(x^*(t), u^*(t), t).\ v(t) + D_2 f(x^*(t), u^*(t), t).\ \delta u(t)$[†]

(17b) $v(0) = \delta x_o$

<u>Proof of Theorem 2.</u> By definition of Fréchet derivative it has to be shown that

(18) $$\lim_{(|\delta x_o| + \|\delta u\|) \to 0} \frac{\|\phi(\cdot, x_o + \delta x_o, u^* + \delta u) - \phi(\cdot, x_o, u^*) - L(\delta x_o, \delta u)\|}{|\delta x_o| + \|\delta u\|} = 0$$

For simplicity in notation let

(19) $$\delta x(t) \triangleq \phi(t, x_o + \delta x_o, u^* + \delta u) - \phi(t, x_o, u^*), \ \forall\, t \in [0,T]$$

In the above notation, using definition of $v(\cdot)$ in (16), (17a) and (17b); (18) reduces to

(20) $$\lim_{(|\delta x_o| + \|\delta u\|) \to 0} \frac{\| \delta x - v \|}{|\delta x_o| + \|\delta u\|} = 0$$

where δx satisfies the d.e.

[†]By virtue of C3] (17a) is a well-defined, linear differential equation.

(21a) $$\delta\dot{x}(t) = f(x^*(t) + \delta x(t),\ u^*(t) + \delta u(t),t) - f(x^*(t),\ u^*(t),t)$$

with

(21b) $$\delta x(0) = \delta x_o$$

We examine (21a) for $t \in [0,T] \sim \mathcal{J}$. By Theorem 1 one can choose $|\delta x_o| + \|\delta u\|$ so small that for a given δ we have that

(22) $$\|\delta u\| + \|\delta x\| < \delta$$

So by mean-value theorem ([9] p 93), (22), and C4]b); for each t in $[0,T] \sim \mathcal{J}$ and each i=1 to n there are real numbers $\lambda_i(t)$ and $\eta_i(t)$ such that

(23) $$0 \leq \lambda_i(t),\ \eta_i(t) \leq 1$$

(24) $$\begin{aligned}\delta\dot{x}_i(t) = {} & D_1 f_i(x^*(t)+\lambda_i(t)\delta x(t),u^*(t)+\eta_i(t)\delta u(t),t).\ \delta x(t) \\ & + D_2 f_i(x^*(t)+\lambda_i(t)\delta x(t),u^*(t)+\eta_i(t)\delta u(t),t).\ \delta u(t) \\ & i = 1,\ldots,\eta\end{aligned}$$

Let the following notation be used to write (24) as a vector equation.

(25) $$\delta\dot{x}(t) = D_1 f_{\lambda,\eta}.\ \delta x(t) + D_2 f_{\lambda,\eta}.\ \delta u(t)$$

We then have for $t \in [0,T] \sim \mathcal{J}$

(26) $$\begin{aligned}\delta\dot{x}(t) - \dot{v}(t) = {} & D_1 f.(\delta x(t) - v(t)) + (D_2 f_{\lambda,\eta} - D_2 f).\ \delta u(t) \\ & + (D_1 f_{\lambda,\eta} - D_1 f)\ \delta x(t)\end{aligned}$$

where we simplified the notation in (17a) and subtracted (17a) from (25) also adding and subtracting the term $D_1 f.\ \delta x(t)$. Letting $I_{\mathcal{J}}$ and $I_{\mathcal{J}^c}$ be the indicator functions of the sets $\mathcal{S}$ and $[0,T] \sim \mathcal{S}$ respectively (26) can be extended to all of the interval [0,T] as follows

$$\delta\dot{x}(t)-\dot{v}(t) = D_1 f . I_{\mathcal{S}^c}(\delta x(t)-v(t)) + (D_2 f_{\lambda,\eta} - D_2 f).I_{\mathcal{S}^c}\delta u(t) + (D_1 f_{\lambda,\eta} - D_1 f).I_{\mathcal{S}^c}\delta x(t) + I_{\mathcal{S}}(\delta\dot{x}(t)-\dot{v}(t)) \tag{27}$$

integrating both sides of (27) and taking norms

$$|\delta x(t)-v(t)| \leq \int_0^t |D_1 f| \; |\delta x(\tau)-v(\tau)| d\tau + \|\delta u\| \int_{[0,T] \sim \mathcal{S}} |D_2 f_{\lambda,\eta} - D_2 f| d\tau + \|\delta x\| \int_{[0,T] \sim \mathcal{S}} |D_1 f_{\lambda,\eta} - D_1 f| d\tau + \int_{\mathcal{S}} (|\delta\dot{x}(\tau)| + \dot{v}(\tau)|) d\tau \tag{28}$$

Using the Bellman-Gronwall Inequality

$$\|\delta x - v\| \leq (I_1 + I_2 + I_3) \exp\left(\int_0^T |D_1 f| d\tau\right) \tag{29}$$

where

$$I_1 \triangleq \|\delta u\| \int_{[0,T] \sim \mathcal{S}} |D_2 f_{\lambda,\eta} - D_2 f| d\tau \tag{30a}$$

$$I_2 \triangleq \|\delta x\| \int_{[0,T] \sim \mathcal{S}} |D_1 f_{\lambda,\eta} - D_1 f| d\tau \tag{30b}$$

$$I_3 \triangleq \int_{\mathcal{S}} (|\delta\dot{x}(\tau)| + |\dot{v}(\tau)|) d\tau \tag{30c}$$

Since the exponential term of inequality (29) is finite by C3] it remains to show that I_1, I_2 and I_3 are 'little o' terms in $|\delta x_o| + \|\delta u\|$. It follows from Theorem 1 that $\|\delta x\|$ divided by $|\delta x_o| + \|\delta u\|$ is bounded above by a constant and obviously $\|\delta u\|$ has the same property the bound being 1. So it suffices, in the cases of I_1 and I_2, to show that the integrals in (30a) and (30b) go to zero as $|\delta x_o| + \|\delta u\|$ goes to zero. Going back to the definitions of $D_1 f_{\lambda,\eta}$ and $D_2 f_{\lambda,\eta}$ using C4]c), making note of the matrix norm given by 4b) a little thought will prove that I_1 and I_2 are 'little o' terms in $|\delta x_o| + \|\delta u\|$.

The next step is to prove the above claim for I_3. For that purpose we make use of the following inequalities which are derivable from earlier results in a straightforward manner.

(31) $$|\delta\dot{x}(t)| \leq k_1(t)\|\delta x\| + k_2(t)\|\delta u\|$$

(32) $$|\dot{v}(t)| \leq |D_1 f|\|v\| + |D_2 f|\|\delta u\|$$

As before $\|\delta x\|$, $\|\delta u\|$ and $\|v\|$† divided by $|\delta x_o| + \|\delta u\|$ are bounded so that it is enough to show that $k_1(\cdot)$, $k_2(\cdot)$, $|D_1 f|$, $|D_2 f|$ integrated on $\mathcal{S}$ can be made arbitrarily small. Since $\mu(\mathcal{S})$ can be made arbitrarily small the result follows from a well known theorem in measure theory ([6] p. 85) using integrability of $k_1(\cdot)$, $k_2(\cdot)$, $|D_1 f|$ and $|D_2 f|$ on [0,T]

5. <u>Conclusions</u>. For a nontrivial and nonsuperficial example of systems to which Theorem 2 applies in a very natural way, the reader is referred to [8]. Simple continuity and compactness arguments show that conditions C1]

---

†It can be easily shown that due to C3] (19) satisfies requirements of Theorem 1 so that $v(\cdot)$ is Lipschitz in $(\delta x_o, \delta u)$. In fact it is linear in $(\delta x_o, \delta u)$.

to C4] are satisfied for differential equations that are smooth in x, u and t.

Extensions of results to the case involving an infinite time interval are by no means trivial due to stability considerations to be taken into account. A treatment of such cases may be found in [10].

References

[1] J. Dieudonné, "Foundations of Modern Analysis," Academic Press, 1960.

[2] E. J. McShane, "Integration," Princeton University Press, 1947.

[3] L. H. Loomis and S. Sternberg, "Advanced Calculus," Addison-Wesley, 1968.

[4] R. E. Kalman, P. L. Falb and M. A. Arbib, "Topics in Mathematical System Theory," Chapter 4.1, McGraw-Hill, 1969.

[5] E. B. Lee and L. Markus, "Foundations of Optimal Control Theory," John Wiley and Sons, 1967.

[6] H. L. Royden, "Real Analysis," MacMillan, 1968.

[7] P. R. Halmos, "Measure Theory," Van Nostrand, 1950.

[8] M. K. Inan and C. A. Desoer, "Optimization of Nonlinear Characteristics," to be published in Journal of Math. Anal. and Appl.

[9] W. Rudin, "Principles of Mathematical Analysis," McGraw-Hill, 1964.

[10] C. A. Desoer and K. K. Wong, "Small Signal Behavior of Nonlinear Networks," Proc. IEEE, 56, 1, pp. 14-23, January 1968.