SOBRE UN PROBLEMA DE SISTEMAS DE ECUACIONES DIFERENCIALES LINEALES

por EMILIO O. ROXIN y VERA W. DE SPINADEL
Instituto de Matemática de la Facultad de Ciencias Exactas y Naturales de Buenos Aires. Dirección Nacional de la Energía Atómica.

RESUMEN. — Consideraremos el comportamiento de las soluciones del sistema de ecuaciones diferenciales lineales con coeficientes variables $dx/dt = A(t)x + f(t)$, donde $x(t)$ es un vector n-dimensional, $A(t)$ una matriz de funciones reales y contínuas de la variable real t y $f(t)$ un vector de componentes $f_i\ (t)$ reales y medibles.

Demostraremos que, dadas las condiciones iniciales, el problema de determinar el vector $f(t)$ tal que haga mínimo el valor de t para el cual $x(t) = 0$, sujeto a la restricción que las componentes $f_i\ (t)$ satisfagan la relación $|f_i\ (t)| \leqq k_i > 0$, admite solución y ésta es tal que $|f_i| = k_i$.

Si el origen es un punto de estabilidad del sistema $dx/dt = A(t)x$ (en el sentido de Liapounoff (1), las constantes k_i pueden elegirse arbitrariamente, por ejemplo, $k_i = 1$. Si, en cambio, el origen es un punto inestable, es posible, dar un criterio para elegir en cada caso las k_i adecuadas para que el problema a que se hace referencia tenga solución.

Este trabajo es una generalización de una memoria reciente de Bellman, Glicksberg y Gross (2), que se refiere al mismo caso para sistemas de ecuaciones diferenciales lineales con coeficientes constantes.

SUMARY. — We shall consider the behaviour of the solutions of the system of linear differential equations with variable coefficients $dx/dt = A(t)x + f(t)$, where $x(t)$ is an n-dimensional vector, $A(t)$ a real, continuous matrix of order n and $f(t)$ a vector whose components $f_i\ (t)$ are real and measurable.

We shall prove that, given the initial conditions, the problem of determining the vector $f(t)$ so as to minimize the value of t required to make $x(t) = 0$, subject to the constraint that the ith. component satisfy the relation $|f_i\ (t)| \leqq k_i > 0$, is solvable and the solution is to take $|f_i| = k_i$.

If the origin is a stable point of the system $dx/dt = A(t)x$ (in the sense of Liapounoff (1)), the constants k_i can be choosed arbitrarily, for example, $k_i = 1$. Furthermore, if the origin is an unstable point, it is possible to give a criterium to choose the adequate k_i to solve the problem.

The present work is a generalization of a recent paper by Bellman, Glicksberg and Gross (2).

1. *Introducción.*

Sea $x(t)$ un vector n-dimensional que satisface a la ecuación diferencial lineal con coeficientes variables

$$\frac{dx(t)}{dt} = A(t)\,x + f(t), \qquad x(0) = x_0, \tag{1.1}$$

donde suponemos que:

a) $A(t)$ es una matriz real y continua, de orden n;
b) $f(t)$ es un vector real y medible cuyas componentes f_i cumplen la relación

$$|f_i| \leqq k_i > 0, \qquad k_i = \text{constantes}. \tag{1.2}$$

Si consideramos el sistema de ecuaciones homogéneo

$$\frac{dx(t)}{dt} = A(t)\,x \tag{1.3}$$

como la descripción del comportamiento de un sistema físico no perturbado, la introducción del término $f(t)$ en la ecuación (1.1) puede interpretarse como la aplicación de una excitación exterior o control. Un problema de importancia a ese respecto es el siguiente: dadas ciertas condiciones iniciales $x(0) = x_0$, determinar el vector $f(t)$ que lleva el sistema físico al estado de reposo en un tiempo t mínimo, imponiendo la condición (1.2) de acotación de las componentes.

En este trabajo demostramos que este problema siempre tiene solución y que el vector $f(t)$ óptimo es tal que

$$f_i(t) = \pm k_i.$$

Se dice en ese caso, que el control es del tipo «on-off» o «bang-bang» (2).

Vamos a distinguir dos posibilidades:

c_1) el origen es un punto *estable* (en el sentido de Liapounoff) para el sistema de ecuaciones (1.3); es decir, dando un $\varepsilon > 0$ existe un $\delta > 0$ tal que toda solución $x_1(t)$ que satisface a la desigualdad

$$|x_1(0)| < \delta,$$

satisface también a

$$|x_1(t)| < \varepsilon;$$

c_2) el origen es un punto *inestable* (en el sentido de Liapounoff) para el sistema de ecuaciones (1. 3).

Veremos que en el primer caso, las constantes $k_i > 0$ pueden elegirse arbitrariamente, por ejemplo, $k_i = 1$. En cambio, en el segundo, la acotación (1. 2) no puede hacerse con constantes arbitrarias, pero en cada caso particular puede hallarse un valor adecuado para ellas, que haga posible la solución del problema arriba mencionado.

2. *Teorema 1.*

Dado el sistema (1. 1) que satisface a las condiciones a), b) y c_1), con el valor de $k_i = 1$, existe un vector $f(t)$ que reduce $x(t)$ a cero con un valor mínimo de t, y para él resulta $f_i(t) = \pm 1$.

Demostración. En virtud de un teorema clásico (3), existe una única solución $x(t)$ del sistema (1. 1) para la que

$$x(0) = x_0.$$

Sea $\phi(t)$ la matriz *resolvente* del sistema (1. 3), o sea la la matriz cuyas n columnas son n soluciones linealmente independientes del sistema (1. 3), tal que

$$\phi(0) = E,$$

donde E es la matriz unidad. $\phi(t)$ resulta real por serlo también $A(t)$. Entonces la solución $x(t)$ está dada por la expresión

$$x(t) = \varphi(t) + \phi(t) \int_0^t \phi^{-1}(s)\, f(s)\, ds, \tag{2.1}$$

siendo $\varphi(t)$ la solución del sistema homogéneo (1. 3) que satisface a

$$\varphi(0) = x_0.$$

A su vez, la $\varphi(t)$ es expresable en la forma

$$\varphi = \phi\, x_0,$$

con lo cual la ecuación (2. 1) se transforma en

$$(2.2) \qquad x(t) = \phi(t) \left[x_0 + \int_0^t \phi^{-1}(s)\, f(s)\, ds \right].$$

El problema se reduce ahora a hallar el vector $f(t)$ que con un valor mínimo de $t = T$ cumpla la relación

$$(2.3) \qquad -x_{0i} = \int_0^T \phi^{-1}(t)\, f(t)\, dt;$$

o bien, elegir las $f_i(t)$ tales que verifiquen

$$(2.4) \qquad -x_{0i} = \int_0^T \sum_{j=0}^{n} \alpha_{ij}(t)\, f_j(t)\, dt; \qquad i = 1, 2, 3, \ldots, n,$$

donde las α_{ij} son los elementos de ϕ^{-1}.

Para ello, vamos a comenzar por demostrar que, dado un valor inicial x_0, existe un vector $f(t)$ y un valor $t = T > 0$ finito tal que vale la ecuación (2. 3). En efecto, siempre nos será posible elegir el vector $f(t)$ tal que

$$\phi^{-1}(t)\, f(t) = K = \text{constante},$$

pues para esto basta tomar

$$(2.5) \qquad f(t) = \phi(t)\, K.$$

Como el sistema homogéneo (1. 3) es estable, $\phi(t)$ es una matriz continua cuya norma $|\phi(t)|$ está acotada para $t > 0$, donde por norma de una matriz entendemos la suma de los valores absolutos de sus elementos. Tomando el vector constante K suficientemente pequeño, podemos siempre hacer que $f(t)$ se mantenga acotada por una constante arbitraria, por ejemplo $|f_i(t)| \leqq 1$ para $t > 0$, en virtud de la desigualdad

$$|f(t)| \leqq |\phi(t)| \cdot |K|.$$

Introduciendo (2. 5) en (2. 3) obtenemos

$$(2.6) \qquad -x_0 = \int_0^T K\,dt = KT,$$

ecuación que determina K y con él, el vector $f(t)$ y el valor $t = T$ tales que se cumplen las ecuaciones (2. 3) y (2. 4).

Veamos ahora qué condiciones debe cumplir el vector $f(t)$ para minimizar el valor de T que reduce $x(T)$ a cero. Para ello seguiremos el razonamiento expuesto en la memoria de Bellman, Glicksberg y Gross (2), que se refiere al caso de ecuaciones diferenciales lineales con coeficientes constantes.

La integral

$$\int_0^T \sum_{j-1}^n \alpha_{ij}(t)\, f_j(t)\, dt$$

define una transformación lineal del espacio de los vectores $f(t)$ en el espacio vectorial n-dimensional de vectores ξ cuya i-ésima componente es

$$\xi_i = \int_0^T \sum_j \alpha_{ij}(t)\, f_j(t)\, dt.$$

Es fácil comprobar que el conjunto convexo de los vectores $f(t)$ sujetos a la condición de acotación $|f_i| \leqq 1$ se transforma en un subconjunto convexo $W(T)$ del espacio euclidiano n-dimensional. Veamos que el conjunto $W(T)$ es creciente con T. En efecto, todo elemento $\xi = \omega_T \bar{f}$ del conjunto $W(T)$ también pertenece al conjunto $W(T')$ donde $T' > T$, ya que la función $\bar{\bar{f}}(t)$ definida del siguiente modo:

$$\bar{\bar{f}}(t) = \begin{cases} \bar{f}(t) & \text{para} \quad t \leqq T \\ 0 & \text{para} \quad t > T \end{cases}$$

tiene por transformada $\omega_{T'}\overline{\overline{f}} = \omega_T \overline{\overline{f}} = \omega_T \overline{f}$, de donde se deduce que $W(T) \subset W(T')$.

El valor de T mínimo buscado es el menor $T < \infty$ para el que $W(T)$ contiene al vector $-x_0$. Como el conjunto $W(T)$ es creciente con T, existe un intervalo (T_0, ∞) para el que $W(T)$ contiene a $-x_0$, mientras que para $T < T_0$, esto no sucede. Si introducimos la métrica euclidiana en el espacio vectorial de los ξ, se puede topologizar el espacio de Banach de los vectores $f(t)$ de manera que la transformación $\omega_{T'}$ resulte continua y el conjunto de los vectores $f(t)$ con $|f_i(t)| \leqq 1$, compacto (4). En ese caso, el conjunto $W(T)$, como imagen continua de un conjunto compacto, es compacto y por consiguiente cerrado, de donde resulta que el conjunto $W(T_0)$ contiene a $-x_0$.

Para $T < T_0$, el vector $-x_0$ no está en $W(T)$, y como el conjunto de los $f(t)$ es convexo, siempre podremos elegir para cada T un vector unitario ϑ_T tal que la proyección de cualquier vector de $W(T)$ sobre él sea menor que la proyección de $-x_0$ sobre el mismo, o sea

$$[\vartheta_T, \omega_T f] \leqq [\vartheta_T, -x_0].$$

Como el conjunto de los vectores de norma unitaria es compacto en la topología euclidiana, existe una sucesión T_n que converge a T_0, para la cual ϑ_{T_n} converge a cierto vector ϑ. Siendo continua la transformación: $\omega_{T_n} f \longrightarrow \omega_{T_0} f$, de donde resulta

$$[\vartheta, \omega_{T_0} f] = \lim_{T_n \to T_0} [\vartheta_{T_n}, \omega_{T_n} f] \leqq \lim_{T_n \to T_0} [\vartheta_{T_n}, -x_0] = [\vartheta, -x_0];$$

es decir, si $f^*(t)$ es un vector para el que

$$\omega_{T_0} f^* = -x_0,$$

se tendrá

$$[\vartheta, \omega_{T_0} f] \leqq [\vartheta, \omega_{T_0} f^*]$$

para todo $f(t)$. Podemos, por tanto, afirmar que existe un conjunto de n constantes ϑ_i no todas nulas para las cuales f^* hace máxima la expresión

$$\sum_i \vartheta_i \int_0^T \sum_j a_{ij}(t)\, f_j(t)\, dt = \sum_j \int_0^T \sum_i \vartheta_i\, a_{ij}(t)\, f_j(t)\, dt.$$

Evidentemente, el máximo de esta expresión es

$$\sum_j \int_0^T \left| \sum_i \vartheta_i \, \alpha_{ij}(t) \right| dt,$$

que se obtiene tomando

$$f_j(t) = \mathrm{sg}\left(\sum \vartheta_i \, \alpha_{ij}(t)\right). \tag{2.7}$$

Hemos así demostrado que si el vector $f(t)$ reduce la solución $x(t)$ a cero para un valor mínimo $t = T$, sus componentes f_i satisfacen a la relación (2.7), de donde se deduce que

$$f_i = \pm 1. \tag{2.8}$$

3. *Teorema* 2.

Dado el sistema (1.1) que satisface a las condiciones a), b) y c_2), siendo k_i constantes sujetas a la acotación

$$k_i = \max_{0 \leq t \leq T} |\phi(t)| \; \frac{|x_0|}{T},$$

donde $\phi(t)$ es la matriz resolvente del sistema (1.3), $T > 0$ arbitrario, existe un vector $f(t)$ que reduce $x(t)$ a cero para un valor mínimo de $t \leqq T$, y para él resulta $|f_i(t)| = k_i$.

Demostración. La demostración es semejante a la del teorema anterior, sólo que, como ahora el sistema (1.3) es inestable, la matriz $\phi(t)$ es tal que su norma $|\phi(t)| \longrightarrow \infty$ para $T \longrightarrow \infty$. Ello trae como consecuencia que el vector $f(t)$ definido por la ecuación (2.5) no permanece acotado. Sin embargo, para cada $T > 0$ podemos tomar un vector constante

$$K = \frac{-x_0}{T}, \tag{3.1}$$

que introducido en la ecuación (2.5) nos da

$$|f(t)| \leqq |\phi(t)| \, . \, |K| = |\phi(t)| \, . \frac{|x_0|}{T}, \tag{3.2}$$

y en el intervalo $0 \leqq t \leqq T$, el vector $f(t)$ cumple la condición

$$|f(t)| \leqq \max_{0 \leq t \leq T} |\Phi(t)| \cdot \frac{|x_0|}{T};$$

o sea

$$|f_i(t)| \leqq k_i.$$

Ese vector $f(t)$ es tal que la solución $x(t)$ se anula para $t = T$, pues de las ecuaciones (3. 1), (2. 5) y (2. 2) resulta

$$x(T) = \Phi(T) \left[x_0 + \int_0^T \Phi^{-1}(t) f(t)\, dt \right] = 0.$$

Hemos así demostrado la existencia de un vector $f(t)$ que hace $x(t) = 0$ para $t = T$. El resto de la demostración sigue las líneas de la del teorema anterior, llegándose a la conclusión que las condiciones que deben cumplir las componentes $f_i(t)$ para minimizar el valor de T, son

$$|f_i(t)| = k_i.$$

Nota sobre el Teorema 2. Con respecto a la acotación impuesta a las constantes k_i en el enunciado de este teorema, cabe señalar que la función $g(T)$ definida por

$$g(T) = \max_{0 \leq t \leq T} |\Phi(t)| \cdot \frac{|x_0|}{T}$$

es una función continua positiva para $T > 0$, que tiende a infinito para $T \to 0$ y cuyo comportamiento para $T \to \infty$ no se puede predecir en general. Esta función posee, por lo tanto, un extremo inferior M y en cada caso podemos elegir un valor T_0 de T tal que

$$g(T_0) < M + \varepsilon$$

siendo $\varepsilon > 0$ arbitrario. La acotación

$$|f_i(t)| \leqq k_i = M + \varepsilon$$

es, a menos del ε, la máxima restricción que, en virtud de este teorema, se puede imponer a las $f_i(t)$.

BIBLIOGRAFIA

(1) LIAPOUNOFF, A. M., Problème Général de la stabilité du mouvement, Princeton University Press, 1947.

(2) BELLMAN, GLICKSBERG and GROSS, On the Bang-Bang Control Problem, Quart. Appl. Math., *XIV*, 1, Abril 1956.

(3) CODDINGTON and LEVINSON, Theory of Ordinary Differential Equations, Mc Graw Hill, 1955.

(4) ALAOGLU, L., Weak topologies of normed linear spaces, Ann. of Math., *41*, 252 - 267, 1940.

ASOCIACION FISICA ARGENTINA

TRIGESIMA PRIMERA REUNION

BUENOS AIRES, Facultad de Ciencias Exactas y Naturales, 23 y 24 mayo de 1958

Informes

"Discusión de experimentos recientes que establecen el significado de los vectores B y H del electromagnetismo clásico", por J. G. Roederer (Comisión Nacional de la Energía Atómica y Facultad de Ciencias Exactas y Naturales). "Análisis de las posibilidades de formación de planetas a consecuencia de explosiones estelares", por F. Cernuschi y S. Codina (Facultad de Humanidades y Ciencias, Montevideo).

Comunicaciones

1. J. STARICCO (Facultad Ing. Bs. As.). *Aplicaciones de la integral multiplicativa.*

Se aplica el concepto de integral multiplicativa o producto integral a la resolución del problema de Cauchy de las ecuaciones diferenciales clásicas en Física y se muestra cómo ciertas formas de operar en Electrodinámica cuántica pueden interpretarse mediante la teoría de dichas integrales.