SEMISISTEMAS DINAMICOS DISCRETOS Y
ALGORITMOS DE MINIMIZACION

por

GERARDO RODRÍGUEZ

Departamento de Análisis Matemático
Universidad de Santiago de Compostela

Considerada una función:

$$\psi : D \subset R^n \to R^1$$

admitiendo un único mínimo x^*, el estudio de la convergencia de ciertos algoritmos de minimización de tipo iterativo, puede ser reducido a un problema de estabilidad asintótica en un conveniente semisistema dinámico discreto sin unicidad (SSDD.).

Uu primer trabajo en este campo es:

«An abstract formulation of minimization algorithms», G. P. SZEGÖ, G. TRECCANI (véase [3]).

En él se presenta la teoría de semisistemas dinámicos discretos sin unicidad y, mediante ella, se prueba, sin hipótesis de convexidad para la función ψ, la convergencia de los algoritmos del gradiente y de las direcciones fijadas.

Posteriormente G. TRECCANI en «A new axiomatization of minimization algorithms» (véase [6]) da un método para construir un conveniente SSDD., para el particular algoritmo de FLETCHER-REEVES con «reset» («reanudación») cada m pasos.

En el presente trabajo se introduce una nueva axiomatización que permite probar la convergencia para una más amplia familia de algoritmos y en la cuál quedan incluidos de manera trivial todos los del gradiente conjugado con «reset».

Las principales ventajas de ella radican sobre todo en:

a) Como se ha dicho, su mayor generalidad.

b) Posibilidad de, mediante ella, obtener nuevos algoritmos ciertamente convergentes.

c) La minimización monodimensional adoptada para la determinación del paso es, al menos bajo ciertas hipótesis acerca de ψ, bastante facilmente realizable desde el punto de vista práctico.

1. — Conceptos topológicos y nociones previas.

Sea X un espacio métrico localmente compacto y $F(X)$ la familia de sus compactos no vacíos. Sea la aplicación:

$\beta : F(X) \times F(X) \rightarrow R^+$ definida por: $\beta(A, B) = \operatorname{Sup} \varrho(x, B) \quad x \in A$

Entonces, si:

$$S_\beta(A, r) = \{B \in F(X) \mid \beta(B, A) < r\}$$

la familia:

$$\mathcal{B} = \{S_\beta(A, r) \mid A \in F(X),\ r > 0\}$$

constituye la base de una topología. Al correspondiente espacio topológico le denotaremos por $F_\beta(X)$. Es de importancia tener en cuenta que si:

$$X_\varrho \equiv (\widetilde{X}, \varrho), \quad X_d \equiv (\widetilde{X}, d)$$

donde ϱ y d son métricas equivalentes, se verifica:

$$F_\beta(X_\varrho) \equiv F_\beta(X_d)$$

Definición 1. — Se denomina semisistema dinámico discreto sin unicidad sobre X a la terna (X, I^+, f) donde:

$$f : X \times I^+ \rightarrow F(X)$$

verifica:

i) $f(x, 0) = \{x\}$

ii) $f(f(x. k), h) = f(x, k + h)$, $\forall x \in X$, $\forall h, k \in I^+$.

iii) $f_k : x \in X \to f_k(x) = f(x, k) \in F_\beta(X)$ es, para todo $k \in I^+$, continua.

Definicion 2. — Se dice que $\chi : J \to X$ es una solución del SSDD. por x si:

i) $J \supset I^+$.

ii) $\chi(i) \in f(\chi(k), i - k)$, $i \geqslant k$.

iii) $\chi(0) = x$.

En lo referente a los conceptos de conjunto límite, atracción, estabilidad asintótica etc., se remite al trabajo ya citado:

«An abstract formulation of minimization algorithms».

Se ha comprobado la verificación de los clásicos teoremas de sistemas dinámicos que caracterizan, mediante funciones de Liapunov la estabilidad asintótica. Con demostraciones que sufren pocas variaciones respecto de las ya conocidas para este tipo de teoremas en sistemas dinámicos, se ha obtenido:

TEOREMA 1. — Sea $M \subset X$ compacto y ϕ una función continua definida en un entorno V de M positivamente invariante, tal que:

i) $\phi(x) = 0$ si $x \in M$, $\phi(x) > 0$ si $x \in V - M$.

ii) $\phi(\chi(1)) < \phi(x)$ para $x \in V - M$ y cualquiera χ por x.

iii) Para todo entorno U de M contenido en V existe $\beta > 0$ tal que $\phi(x) > \beta$, $\forall x \in V - M$.

Entonces, M es asintóticamente estable.

TEOREMA 2. — Sea M asintóticamente estable, con región de atracción $A(M)$. Entonces, existe una función:

$$\phi : A(M) \to R^1$$

semicontinua superiormente, verificando las condiciones i), ii), iii) del teorema 1, en el entorno $V \equiv A(M)$.

Sea ahora $\psi : D \subset R^n \rightarrow R^1$ una función de clase uno, tal que:

i) Admite un conjunto de nivel compacto L.

ii) Existe en L un único punto crítico x^*, del que se sabe que es mínimo.

Nuestro objetivo es la construcción de un conveniente semisistema dinámico discreto sin unicidad, que permita probar la convergencia de cierta familia de algoritmos de minimización para una función en las anteriores condiciones.

2. — Construcción del SSDD. Estabilidad asintótica.

Dado un espacio métrico X y una aplicación:

$$f(\cdot, 1) : X \rightarrow F_\beta (X)$$

continua, es posible construir un SSDD. adoptando como:

$$f : X \times I_+ \rightarrow F_\beta (X)$$

la definida por:

$$f(x, k+1) = f(f(x, k), 1)$$

Se comenzará por la construcción del espacio métrico adecuado. Sea $\widetilde{L}$ el soporte del subespacio métrico de R^n, L, y $\widetilde{A}_i = \underbrace{\widetilde{L} \times \ldots \times \widetilde{L}}_{i}$. Fijado $m \in I^+$, considérese el conjunto:

$$\widetilde{A} = \bigcup_{i=1}^{m} \widetilde{A}_i$$

Entonces, si ϱ_i es una métrica definida en $\widetilde{A}_i$, acotada por uno, y equivalente a la del espacio métrico $\underbrace{L \times \ldots \times L}_{i}$, se verifica:

Teorema 3. — La aplicación $\varrho : \widetilde{A} \times \widetilde{A} \rightarrow R^1$, definida por:

$$\varrho (x, y) = \begin{cases} \varrho_i (x, y) & \text{si} \quad x, y \in \widetilde{A}_i \\ 1 & \text{si} \quad x \in \widetilde{A}_i, \; y \in \widetilde{A}_j, \; i \neq j \end{cases}$$

es una métrica y el correspondiente espacio, localmente compacto.

El espacio métrico en el que se trabajará será el:

$$X = \{\bar{x} = (x_1, ..., x_s) \in A, \ s \leqslant m \mid \psi(x_{i+1}) \leqslant \psi(x_i), \ i \leqslant s - 1\}$$

que evidentemente es, también, localmente compacto. Si X_i es el subespacio métrico de X que tiene como conjucto base $\widetilde{A}_i$ se puede escribir:

$$X = X_1 \, U \, ... \, UX_m.$$

A la vista del particular espacio métrico utilizado, la construcción de la antedicha aplicación $f(\cdot, 1)$ se consigue con la de una:

$$\widetilde{f}(\cdot, 1) : X_1 \, U \, ... \, UX_{m-1} \to F_\beta(X)$$

continua. En efecto:

Para $\bar{x} = (x_1, ..., x_s)$, $s < m$ y $k \leqslant m - s$, sea $\widetilde{f}(\bar{x}, k) = f(f(\bar{x}, k-1), 1)$.

Entonces, basta con adoptar:

$$f(\bar{x}, 1) = \begin{cases} \widetilde{f}(\bar{x}, 1) & \text{si} \quad \bar{x} = (x, ..., x_s) \quad \text{con} \quad s < m \\ \widetilde{f}(x_m, m-1) & \text{si} \quad \bar{x} = (x_1, ..., x_m) \end{cases}$$

que como se prueba trivialmente, resuelve el problema.

Sea ahora H' la esfera de R^n de centro $\theta = (0, ..., 0)$ radio 1. Sea $H = H' \, U \, \{\theta\}$.

Definicion 3. — Se denomina aplicación de direcciones en el espacio métrico X, a toda aplicación:

$$\mathcal{F} : X_1 \, U \, ... \, UX_{m-1} \to F_\beta(H)$$

continua.

Suponiendo, en lo sucesivo que una aplicación de direcciones ha sido dada, se procederá a la construcción de la aplicación $\widetilde{f}(\cdot, 1)$. Como elementos auxiliares, fíjense $H > 0$ y $t \in (0, 1)$ arbitrarios.

Dados $\bar{x} = (x_1, ..., x_s)$, $s < m$, y $p \in \mathcal{F}(\bar{x})$, será utilizada la siguiente notación:

$$G_p = G(\bar{x}, p) = \{\alpha \in R \mid x_s + \alpha p \in \mathring{L}\}$$

$$\bar{\alpha}(x, p) = \bar{\alpha}_p = \operatorname{Inf}\left\{\gamma \in [0, H] \;\middle|\; \begin{array}{l} \alpha \in G \to \psi(x_s + \alpha p) \text{ presenta en } \gamma \\ \text{punto crítico o mínimo local en } [0, H] \end{array}\right\}$$

$$\eta_p = t\,\bar{\alpha}_p, \quad I_p = [\eta_p, H]$$

Entonces, la aplicación:

$$\widetilde{f}(\cdot\,, 1) : X_1 U \ldots U X_{m-1} \to F_\beta (X)$$

es la definida por:

$$\widetilde{f}(\bar{x}, 1) = \{(x_1, \ldots, x_s, y) \mid y = x_s + \alpha \cdot p,\ p \in \mathcal{F}(\bar{x})\}$$

siendo:

$$\alpha \in I_p \quad \text{con} \quad \psi(y) \leqslant \psi(x_s + \eta_p \cdot p)$$

TEOREMA 4. — La aplicación $\widetilde{f}(\cdot\,, 1)$ es continua.

Demostracion: Si el teorema no fuese cierto en $\bar{x}$, existirían $\varepsilon > 0$ y $\{\bar{x}^n\}$ tal que:

$$\{\bar{x}^n\} \to \bar{x} \text{ y } \beta(\widetilde{f}(\bar{x}^n, 1), \widetilde{f}(\bar{x}, 1)) \geqslant \varepsilon$$

Sea $\{\bar{z}^n\}$, $\bar{z}^n \in \widetilde{f}(\bar{x}^n, 1)$, tal que $\varrho(\bar{z}^n, \widetilde{f}(\bar{z}, 1)) \geqslant \varepsilon$

Se denotará:

$$\bar{x}^n = (x_1^n, \ldots, x_s^n),\ \bar{x} = (x_1, \ldots, x_s),\ \bar{z}^n = (x_1^n, \ldots, x_s^n, z^n)$$
$$A = \{z \mid (x_1, \ldots, x_s, z) \in \widetilde{f}(\bar{x}, 1)\}$$

y para la sucesión $\{z^n\}$, últimamente, se tiene: $\varrho(z^n, A) \geqslant \varepsilon$,
Por construcción:

$$z^n = x_s^n + \alpha^n \cdot p^n \quad \text{con} \quad p^n \in \mathcal{F}(\bar{x}^n), \quad \text{y con} \quad \alpha^n \in [\eta_n, H]$$

verificandose:

$$\psi(x_s^n + \alpha^n \cdot p^n) \leqslant \psi(x_s^n + \eta_n \cdot p^n). \text{ (Se ha escrito } \eta_n = t\,\bar{\alpha}(x^n, p^n))$$

Como $\{\mathcal{F}(\bar{x}^n)\} \xrightarrow{\beta} \mathcal{F}(\bar{x})$ y $[0, H]$ es compacto, se puede suponer:

$$\{p^n\} \to \widetilde{p} \in \mathcal{F}(\bar{x}), \quad \{\alpha^n\} \to \widetilde{\alpha} \in [0, H]$$

Mediante un cálculo bastante largo se demuestra que:

$$(x_1, \ldots, x_s, x_s + \widetilde{\alpha} \cdot \widetilde{p}) \in \widetilde{f}(\bar{x}, 1)$$

lo que está en contra de lo supuesto pues:

$$\{(x_1^n, \ldots, x_s^n, x_s^n + \alpha^n \cdot p^n)\} \to (x_1, \ldots, x_s, x_s + \tilde{\alpha}\,\tilde{p})$$

Al objeto de estudiar la estabilidad asintótica, hay que considerar una función de Liapunov. Se adopta la:

$\Psi : X \to R^1$, definida por: $\Psi(\bar{x}) = \psi(x_s) = \text{Min}\ \{\psi(x_i) \mid i = 1, \ldots, s\}$

Definicion 4. — La aplicación de direcciones $\mathcal{F}$ y, consecuentemente, el semisistema, se dice descendente para ψ cuando para todo $x \in X_1$ se verifica:

$$\bar{y} \in \tilde{f}(x, 1) \Rightarrow [\Psi(\bar{y}) < \psi(x) \Leftrightarrow x \neq x^*]$$

TEOREMA 5. — El compacto $M = \{(x_1, \ldots, x_m) \in X \mid x_m = x^*\}$ es globalmente asintóticamente estable.

Demostración: Basta con tener en cuenta que la función:

$$\varphi : X_m \to R^1 \quad \text{definida por:} \quad \Psi(\bar{x}) = \psi(\bar{x}) - \psi(x^*)$$

verifica las condiciones del teorema 1.

3. — APLICACIÓN A ALGORITMOS DE MINIMIZACIÓN.

Se utilizará la construcción anterior para probar la convergencia de un tipo de algoritmos de minimización.

Hipótesis acerca de la función: Las enunciadas para la función ψ (Nótese que no se hacen hipótesis acerca de su convexidad).

Algoritmo: Considérese el algoritmo de minimización siguiente:

$$x^{k+1} = x^k + \alpha_k \cdot p_N^k \quad \text{siendo:} \begin{cases} p^0 = -g^0 \\ p^{k+1} = -g^{k+1} + \beta_k p^k \end{cases}$$

y donde se ha denotado: $p_N^k = \dfrac{p^k}{\|p^k\|}$, $g^i = \text{grad}\ \psi(x^i)$

y donde β es una función real, continua, de $x^0, x^1, \ldots, x^{k+1}$. Cuando se obtiene el punto x^{m-1}, se detiene el anterior proceso y se inicia de nuevo partiendo de x^{m-1} como punto inicial. Si algún $p^k = \theta$, se inicia el algoritmo en x^k.

Según la elección de los números β_k, el algoritmo recibe nombres diversos. Las elecciones más corrientes son:

I) $$\beta_k = \frac{\langle g^{k+1}, Gp^k \rangle}{\langle p^k, Gp^k \rangle}$$

Siendo G una matriz simétrica definida positiva y $\langle \cdot\, , \cdot \rangle$ el producto interior en R^n.

II) $\beta_k = \dfrac{\|g^{k+1}\|}{\|g^k\|}$ (Algoritmo de FLETCHER-REEVES).

III) $\beta_k = \dfrac{\langle g^{k+1} - g^k, g^k \rangle}{\|g^k\|}$ (Algoritmo de POLAK-RIBIERE).

Elección del paso α_k. — Fígense arbitrariamente $H > 0$ y $t \in (0, 1)$. Considerada la función:

$$\varphi : \beta \in G_k \subset R^1 \to \varphi(\beta) = \psi(x^k + \beta p_N^k) \in R^1,$$

se distinguen dos posibilidades:

1.º) Si φ admite en $\beta = 0$ mínimo local o es creciente en dicho punto, se adopta $\alpha_k = 0$.

2.º) Supóngase que no estamos en el caso anterior. Determínese $\gamma < H$ tal que en $[0, \gamma]$ no presente la anterior aplicación mínimo local (podría existir en $[0, \gamma]$ punto de inflexión). Desde luego, φ es decreciente en $[0, \gamma]$.

Sea ahora $\eta = \gamma - t\gamma$ y $\mathcal{P} = \{\beta_0, ..., \beta_s\}$ una partición de $[\gamma, H]$ de norma menor o igual a η. Tómese el subconjunto de $\mathcal{P}$,

$$\mathcal{P}' = \{\beta_0, ..., \beta_h\}, \qquad h \leqslant s,$$

de la siguiente manera:

a) Puede ser elegido $\mathcal{P}' = \mathcal{P}$, es decir, $h = s$.

b) Se podrá elegir $\mathcal{P}' \neq \mathcal{P}$, es decir $h < s$, si y sólo si existe algún criterio que permita afirmar que en $[0, \beta_h]$ exista punto crítico de φ.

Sea $\bar{\beta} = \beta_r = $ Máx {mínimos absolutos de φ en $\mathcal{P}'$} y:

$$\widehat{\beta} = \begin{cases} \text{Arg Min } \{\varphi(\beta) \mid \beta = \beta_{r-1}, \beta_{r+1}\} & \text{si} \quad r > 0 \\ \beta_1 \quad \text{si} \quad r = 0 \end{cases}$$

Considerada la parábola pasando por:

$$(x^k, \varphi(0)), \qquad (x^k + \bar{\beta} p_N^k, \varphi(\bar{\beta})), \qquad (x^k + \widehat{\beta} p_N^k, \varphi(\widehat{\beta}))$$

supóngase que alcanza su mínimo en $\widetilde{\beta}$. Entonces:

$$\alpha_k = \text{Arg Min } \{\varphi(\beta) \mid \beta = \bar{\beta}, \widetilde{\beta}\}$$

Observación: Podría haberse tomado $\alpha_k = \bar{\beta}$, sin efectuar la interpolación parabólica, y el teorema que sigue sería igualmente cierto.

TEOREMA 6. — Si $\{x^h\}$ es una sucesión generada por el anterior algoritmo, se verifica:

$$\lim_{h \to \infty} \{x^h\} = x^*$$

Demostración: Consiste en construir, a partir de los datos del algoritmo, una aplicación de direcciones, descentes para ψ, adecuada. En el caso de los ejemplos mencionados se procede del siguiente modo:

Sea $\bar{x} = (x^0, \ldots, x^s)$ con $s < m - 1$. Se definen recursivamente (cuando sea posible) los vectores $q^0, q^1, \ldots, q^s$, del siguiente modo:

$$q^0 = - g^0$$

Supóngase que ha sido construido q^{k-1}. Si tiene sentido, en cada caso, el número real:

$$\eta_{k-1} = \frac{\langle g^k, G p^{k-1} \rangle}{\langle p^{k-1}, G p^{k-1} \rangle}, \quad \eta_{k-1} = \frac{\|g^k\|}{\|g^{k-1}\|}, \quad \eta_{k-1} = \frac{\langle g^k - g^{k-1}, g^{k-1} \rangle}{\|g^{k-1}\|}$$

se construye:

$$q^k = - g^k - \eta_{k-1} g^{k-1} - \ldots - \eta_{k-1} \eta_{k-2} \cdots \eta_0 g_0$$

y la aplicación de direcciones $\mathcal{F}$ viene dada por:

$$\mathcal{F}(x) = \begin{cases} q_N^s & \text{si} \quad q^s \text{ está definido y } q^s \neq \theta \\ H & \text{si} \quad q^s \text{ no está definido o } q^s = \theta \end{cases}$$

Entonces, mediante los números H y t utilizados en el algoritmo, se construye el correspondiente semisistema. Para él, la sucesión $\{x^n\}$, proporciona la solución:

$$\chi : I^+ \to X$$

definida por:

$$\chi(0) = x^0, \quad \chi(1) = (x^0, x^1), \ldots, \chi(m-1) = (x^0, \ldots, x^{m-1})$$
$$\chi(m+h) = (x^{(h+1)m-(h+1)}, \ldots, x^{(h+2)m-(h+2)}), \quad h \geqslant 0$$

Por aplicación del teorema 5 concluye la demostración.

BIBLIOGRAFÍA

1. BHATIA N. P.; SZEGÖ G. P. *Stability Theory of Dynamical Systems*. Springer Verlag, Berlín, Heidelberg. New York. 1970.

2. CEA J. *Optimisation theorie et algoritms*, Dunod. París. 1971

3. KUHN M. W.; SZEGÖ G. P. *Differential Games and Related Topics*. North-Holland. 1971.

4. MICHAEL E. *Topologies on Spaces of Subsets*. Trans. Amer. Math. Soc. 71, 152-182. 1951.

5. ORTEGA J. M.; RHEINBOLDT W. C. *Iterative solution of nonlinear equations in several variables*. Academic Press 1970.

6. SZEGÖ G. P. *Minimization Algorithms. Mathematical Theories and Computer Results*. Academic Press. 1972.