

## Acerca de la inmunidad a las perturbaciones en el método de descenso por el gradiente\*

ALINA FEDÓSOVA†

### Resumen

En el presente trabajo se realiza un experimento computacional (basado en resultados teóricamente verificados) acerca del error en la convergencia en el caso del empleo del método del gradiente en la programación no lineal, utilizando dos conocidas funciones de prueba de varias variables y combinando diversas elecciones del paso.

#### 1. Sea dado el problema

$$f(u) \rightarrow \inf, \quad u \in \mathbb{R}^n,$$

suponiendo que  $f$  es continuamente diferenciable en  $\mathbb{R}^n$ . Como se sabe (ver, por ejemplo, [1] ó [2]), en todos los métodos iterativos, en particular en el del gradiente, se supone la elección de un punto  $u_0$  de partida que aproxime la solución, elección para la cual no existen reglas generales. Si ya ha sido elegido ese punto  $u_0$ , el método de descenso por el gradiente consiste en construir una sucesión  $\{u_k\}$  según la fórmula

$$u_{k+1} = u_k - a_k \cdot f'(u_k), \quad a_k > 0 \text{ para } k = 0, 1, 2, \dots \quad (1)$$

El número  $a_k$  se denomina longitud del paso, o simplemente paso, del método. Si  $f'(u_k) \neq 0$  es posible elegir el paso de manera que  $f'(u_{k+1}) < f'(u_k)$ ; si  $f'(u_k) = 0$  el punto  $u_k$  será un punto estacionario, caso en el

\*El presente trabajo se basa en la tesis de grado en matemáticas de la autora, presentada en la Universidad Estatal de Moscú en 1990.

†Docente de la Universidad Autónoma de Bucaramanga, Bucaramanga, COLOMBIA.

cual el proceso se detiene y, si es necesario, se verifica la conducta de la función en los alrededores de  $u_k$ , a fin de establecer si en ese punto se logra o no el mínimo de  $f$ . En particular, si  $f$  es convexa el punto  $u_k$  siempre será minimal.

Existen diferentes métodos para la elección de los pasos  $a_k$  en la fórmula (1). Dependiendo de esos métodos se pueden obtener diferentes variantes del método de descenso por el gradiente. Aquí consideramos cuatro maneras de elegir el paso:

- (1) Paso constante. Si  $f \in C^{1,1}(\mathbb{R}^n)$ , es decir, si  $f \in C^1(\mathbb{R}^n)$  y el gradiente  $f'$  satisface la condición

$$|f'(u) - f'(v)| \leq L|u - v|, \quad u, v \in \mathbb{R}^n,$$

en donde  $L$  se conoce, en calidad de  $a_k$  se puede tomar, en virtud de (1), cualquier número que satisfaga la condición

$$0 < \varepsilon_0 \leq a_k \leq \frac{2}{L + 2\varepsilon},$$

en donde  $\varepsilon_0$  y  $\varepsilon$  son parámetros positivos del método. En particular, con  $\varepsilon = 1/2$  y  $\varepsilon_0 = 1/L$  se obtiene el método (1) con paso constante  $a_k = 1/L$ .

- (2) En segundo lugar examinamos el método de elección del paso que se desprende de la llamada condición de Armijo [3]:

$$f(u_k) - f(u_k - a_k) \cdot f'(u_k) \geq \varepsilon a_k |f'(u_k)|^2, \quad \varepsilon > 0.$$

Por lo general se toma, para satisfacer esa condición, un  $a_k = a > 0$  (el mismo en todas las iteraciones; en el programa utilizado se toma  $a = 1$ ). Luego, si es necesario, el  $a_k$  se parte, es decir, se varía según la fórmula  $a_k = \lambda^i a$ ,  $i = 0, 1, 2, \dots$ ,  $0 < \lambda < 1$ , mientras no se cumpla por primera vez la condición de Armijo.

- (3) El método del descenso más rápido. Sobre el rayo

$$\{u \in \mathbb{R}^n : u = u_k - a \cdot f'(u_k), \quad a \geq 0\}$$

y en la dirección del antigradiente, se introduce la función de una variable

$$f_k(a) = f(u_k - a \cdot f'(u_k)), \quad a \geq 0,$$

y  $a_k$  se determina a partir de las condiciones

$$f_k(a_k) = \inf_{a \geq 0} f_k(a), \quad a \geq 0.$$

Para obtener aquí el paso optimal es necesario minimizar una función de una variable, por lo cual es conveniente emplear el método de la sección dorada (ver, por ejemplo, [4]). Una vez que el intervalo se hace menor que un  $E$  dado de antemano (por ejemplo 0,01 ó 0,001), se detiene la subdivisión.

- (4) Examinamos por último el método consistente simplemente en elegir  $a$  según la fórmula  $a_k = 1/k$  en donde  $k$  es el número de la correspondiente iteración.

Supongamos, pues, que ya ha sido escogido el método de elección del paso  $a_k$  en la fórmula (1). Consideremos varias maneras de abandonar el ciclo de iteraciones en el método del gradiente. Se puede, en primer lugar, dar un máximo número de iteraciones: una vez el contador del ciclo lo alcance, el programa termina (en consecuencia, si el usuario no desea emplear este criterio, es necesario tomar ese número suficientemente grande). En programas experimentales, y cuando se conoce teóricamente el punto minimal (como sucede en el presente trabajo), se puede parar el programa cuando el valor en una determinada iteración difiere del mínimo teórico en una cantidad prefijada. Finalmente, se puede dar por finalizado el programa cuando el tamaño (norma) del gradiente se reduzca por debajo de algún número dado de antemano.

2. Tomaremos como funciones de prueba las conocidas funciones-test de Rosenbroke (a) y de Powell (b):

$$(a) f : \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto f(x) = 100(x_2 - x_1)^2 + (1 - x_1)^2$$

$$(b) f : \mathbb{R}^4 \rightarrow \mathbb{R}, x \mapsto f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$$

Como se puede ver fácilmente, los puntos minimales son, respectivamente,  $x^* = (1, 1)$  y  $x^* = (0, 0, 0, 0)$ , con  $f(x^*) = 0$  en ambos casos. Examinaremos la efectividad de los métodos descritos desde diferentes puntos de vista: cuál de ellos funciona mejor en las primeras iteraciones, cuál en las finales, cuál es el método más sensible a los errores y cuál es la influencia de las perturbaciones aleatorias.

Las pruebas se efectúan desde diferentes puntos de salida, y con diferentes pasos en la fórmula para la obtención del gradiente. En el programa (en FORTRAN) se utilizó:

$N$  es el número de iteraciones;

$Q = \frac{f(x_n) + DEL}{F(x_{n-1}) + DEL}$ ,  $n = 1, 2, \dots$ , en donde  $DEL$  representa la perturbación;

$\sqrt{G_1^2 + \dots + G_N^2}$  es la norma del gradiente;

$f^* = f(x^*)$ ;

$R1$  y  $R2$  son registros de control, así:

$$R1_n = \begin{cases} R1_{n-1}, & \text{si } f(x^n) \geq f(x^{n-1}), \\ f(x^n) & \text{en caso contrario;} \end{cases}$$

$$R2_n = \begin{cases} R2_n, & \text{si } f(x^n) + \Delta_0(x^n) \geq f(x^{n-1}) + \Delta_0(x^{n-1}), \\ f(x^n) + \Delta_0(x^n) & \text{en caso contrario.} \end{cases}$$

En las Tablas 1 y 2 se muestran los resultados para la función de Rosenbroke.

**Tabla 1:** Valores de la función de Rosenbroke con paso inicial  $H = 0,005$ ,  $E = 0,01$  y punto inicial  $X = (1,2, 1,0)$ .

Método del paso constante	Condición de Armijo	Descenso más rápido	Paso $1/n$	Cantidad de iteraciones
3,9146388	0,0398397	3,4513121	4,2427511	100
3,5170040	0,0398397	2,4488373	4,2373786	200
3,1216450	0,0398397	0,6770810	4,2342138	300
2,7250426	0,0398397	0,1563617	4,2319355	400
2,3267090	0,0398397	0,0789645	4,2301621	500
1,9246790	0,0398397	0,0433036	4,2287374	600
1,5159315	0,0398397	0,0323619	4,2274528	700
1,1048393	0,0398397	0,0133441	4,2263494	800
0,7199323	0,0398397	0,0072542	4,2253809	900
0,4421650	0,0398397	0,0042056	4,2245221	1000
0,2829309	0,0398397	0,0034611	4,2237344	1100
0,1913872	0,0398397	0,0050416	4,2230177	1200
0,1325745	0,0398397	0,0075271	4,2223549	1300
0,0922595	0,0398397	0,0119325	4,2217417	1400
0,0633566	0,0398397	0,0181235	4,2211742	1500
0,0426308	0,0398397	0,0248027	4,2206292	1600
0,0275411	0,0398397	0,0334600	4,2201302	1700
0,1656420	0,0398397	0,0436508	4,2196584	1800
0,0090933	0,0398397	0,0551066	4,2192082	1900
0,0052360	0,0398397	0,0673533	4,2190510	2000

En la cuarta columna de la anterior tabla se tomó como punto inicial aquél en el cual se interrumpió la monotonía del primer método:  $(-1.071, 1.145)$  (en la octava iteración). Se nota que en el primer caso la función disminuye muchísimo más rápido que en el cuarto, y que el método con la condición de Armijo llega hasta un cierto punto y deja luego de minimizar la función.

**Tabla 2:** Valores de la función de Rosenbroke con  $H = 0,01$  y punto inicial  $X = (1,2, 1,0)$

Método del paso constante	Condición de Armijo	Descenso más rápido	Paso $1/n$	Cantidad de iteraciones
0,5005036	0,7793	0,1957095	0,0786937	100
0,329859	0,7793	0,0885157	0,078393	200
0,2225903	0,7793	0,0396433	0,0781298	300
0,1495631	0,7793	0,022266	0,0779367	400
0,0973408	0,7793	0,013898	0,0777022	500
0,0607789	0,7793	0,0155612	0,0775403	600
0,0354106	0,7793	0,02250476	0,0774058	700
0,0203441	0,7793	0,0406499	0,0772921	800
0,0138946	0,7793	0,0651653	0,0771914	900
0,015533	0,7793	0,0969704	0,0777616	1000
0,0279076	0,7793	0,137699	0,0769651	1100
0,0487215	0,7793	0,1684848	0,0768725	1200
0,0773981	0,7793	0,1844708	0,0768563	1300
0,1177751	0,7793	0,2009083	0,0768321	1400
0,1680612	0,7793	0,2186577	0,0767957	1500
0,2293799	0,7793	0,236936	0,0767582	1600
0,3007208	0,7793	0,256126	0,0767243	1700
0,3832172	0,7793	0,27637	0,0766936	1800
0,4780249	0,7793	0,2976609	0,076661	1900
0,5845821	0,7793	0,3189981	0,0766312	2000

Ahora miremos los resultados para la función de Powell. Se investigó desde diferentes puntos iniciales y con diferentes pasos.

a) Con punto inicial  $X = (3,0, -1,0, 0,0, 1,0)$  (Tabla 3).

**Tabla 3:** Valores de la función de Powell con  $H = 0,01$ ,  $E = 0,001$  y punto inicial  $X = (3,0, -1,0, 0,0, 1,0)$ .

Método del paso constante	Condición de Armijo	Descenso más rápido	Cantidad de iteraciones
6,9176278	0,989767	0,5976278	100
3,3310902	0,1915855	0,0989767	200
1,7142875	0,1915855	0,393094	300
0,9449605	0,1915855	0,0244672	400
0,5554196	0,1915855	0,0195853	500
0,3492664	0,1915855	0,0173583	600
0,2334643	0,1915855	0,0166871	700
0,1642882	0,1915855	0,0151131	800
0,1212971	0,1915855	0,0145971	900
0,0933054	0,1915855	0,0147061	1000
0,0747128	0,1915855	0,0145480	1100
0,0614184	0,1915855	0,0147101	1200
0,0517934	0,1915855	0,0139693	1300
0,0442324	0,1915855	0,0142718	1400
0,0386819	0,1915855	0,0145048	1500
0,0344183	0,1915855	0,0143259	1600
0,0311954	0,1915855	0,0139033	1700
0,0285636	0,1915855	0,0144576	1800
0,0265812	0,1915855	0,0139366	1900
0,0250686	0,1915855	0,0139323	2000

En la Tabla 3 se nota que en las últimas etapas del cómputo el método del descenso más rápido funciona peor que el de paso constante, mientras que el método con elección del paso según la regla de Armijo se precipita con gran velocidad al punto minimal pero es el menos estable a los errores.

- b) Veamos lo que sucede arrancando del mismo punto, pero reduciendo el paso inicial anterior a la mitad (Tabla 4).

**Tabla 4:** Valores de la función de Powell con  $H = 0,01$ ,  $E = 0,001$  y punto inicial  $X = (3,0, -1,0, 0,0, 1,0)$ .

Método del paso constante	Condición de Armijo	Descenso más rápido	Cantidad de iteraciones
6,3778439	0,0210521	0,574846	100
3,2973967	0,0210394	0,0883952	200
1,6849695	0,0210296	0,0304305	300
0,9183398	0,0210252	0,0167928	400
0,5343847	0,0210252	0,0118957	500
0,3330246	0,0210252	0,0095035	600
0,2184905	0,0210252	0,0085722	700

Método del paso constante	Condición de Armijo	Descenso más rápido	Cantidad de iteraciones
0,1514657	0,0210252	0,0081503	800
0,1099171	0,0210252	0,0074709	900
0,0828499	0,0210252	0,007433	1000
0,0647942	0,0210252	0,0075742	1100
0,0514948	0,0210252	0,007473	1200
0,0422315	0,0210252	0,0070263	1300
0,0351916	0,0210252	0,0065711	1400
0,0298374	0,0210252	0,0066636	1500
0,0254829	0,0210252	0,0064446	1600
0,0226657	0,0210252	0,006691	1700
0,0200892	0,0210252	0,0067772	1800
0,0183458	0,0210252	0,0062654	1900
0,0169645	0,0210252	0,0065286	2000
2,4133768	0,0253318	0,2479901	100
1,0860076	0,0253298	0,0436867	200
0,6035725	0,0253273	0,0133709	300

- c) En el siguiente ensayo se parte del punto  $X = (1.0, -2.0, 3.0, -1.2)$  con paso  $H = 0,005$  (Tabla 5).

Tabla 5: Valores de la función de Powell con  $H = 0,005$ ,  $E = 0,001$  y punto inicial  $X = (1.0, -2.0, 3.0, -1.2)$ .

Método del paso constante	Condición de Armijo	Descenso más rápido	Cantidad de iteraciones
0,3573911	0,0253262	0,0063005	400
0,2233074	0,0253250	0,0036045	500
0,1471507	0,0253239	0,0024612	600
0,1111502	0,0253220	0,0017609	700
0,0720535	0,0253204	0,0009899	800
0,0531655	0,0253193		900
0,0403974	0,0253181		1000
0,0314170	0,0253167		1100
0,0248921	0,0253147		1200
0,0203892	0,0253136		1300
0,0168777	0,0253125		1400
0,0140467	0,0253115		1500
0,0120654	0,0253097		1600
0,0102160	0,0253074		1700
0,0086796	0,0253062		1800
0,0076593	0,0253		1900
0,0069196	0,0253		2000

Como se ve, hacia la iteración 800 el método del mayor descenso alcanza el nivel dado (0,001) y funciona mejor que el método de paso constante (si no se tiene en cuenta el gasto de tiempo para calcular el paso optimal en la fórmula iterativa del método de la sección dorada).

- d) En la Tabla 6 se utiliza de nuevo la función de Powell, pero con punto de salida (3.5, 0, 1.2, 1) y  $H = 0,01$ , en donde se puede ver que el método del mayor descenso funciona mejor al principio, pero después comienza a comportarse mejor el del paso constante.

Tabla 6: Valores de la función de Powell con  $H = 0,01$ ,  $E = 0,001$  y punto inicial  $X = (3.5, 0, 1.2, 1)$ .

Método del paso constante	Condición de Armijo	Descenso más rápido	Cantidad de iteraciones
7,9693108	0,1620509	0,4610387	100
3,802439	0,1620509	0,0872504	200
1,9329621	0,1620509	0,0360324	300
1,0495958	0,1620509	0,0236068	400
0,6106259	0,1620509	0,0185616	500
0,3791308	0,1620509	0,0168383	600
0,2501405	0,1620509	0,0152543	700
0,1742459	0,1620509	0,0158312	800
0,1276352	0,1620509	0,0148460	900
0,0979405	0,1620509	0,0147091	1000
0,0781021	0,1620509	0,0152394	1100
0,0640563	0,1620509	0,0147310	1200
0,0533711	0,1620509	0,0146273	1300
0,0460106	0,1620509	0,0143663	1400
0,0398326	0,1620509	0,0143602	1500
0,0351816	0,1620509	0,0142154	1600
0,0317874	0,1620509		1700
0,0291725	0,1620509		1800
0,0268576	0,1620509		1900
0,0254451	0,1620509		2000

- e) Una variante del método del gradiente, en el cual el paso se calcula por la fórmula  $ALF = I/(500i)$ , en donde  $i$  es el número de la correspondiente iteración, desciende muy lentamente. En las tablas 7 y 8 se hace la comparación de esta variante con el método 1), pero como punto inicial se ha tomado el punto de interrupción de la monotonía del método 1).



**Tabla 7:** Valores de la función de Powell con  $H = 0,005$  y punto inicial  $X = (0,403, -0,043, 0,175, 0,210)$ .

Método del paso constante	Método con paso $I/i$	Cantidad de iteraciones
0,0422315	0,0439423	100
0,0351916	0,0433291	200
0,0298374	0,0437690	300
0,0254829	0,0437365	400
0,0226657	0,0437159	500
0,0200892	0,0437008	600
0,0183458	0,0436875	700
0,0169645	0,0436760	800
0,0158442	0,0436662	900
0,0149124	0,0436574	1000
0,0139479	0,0436493	1100
0,0133213	0,0436417	1200
0,0129526	0,0436352	1300
0,0123457	0,0436290	1400
0,0120055	0,0436232	1500
0,0115338	0,0436180	1600
0,0112485	0,0436130	1700
0,0111550	0,0436082	1800
0,0109048	0,0436037	1900
0,0107824	0,0435997	2000

**Tabla 8:** Valores de la función de Powell con  $H = 0,01$ ,  $E = 0,001$  y punto inicial  $X = (0,429, -0,049, 0,1815, 0,227)$ .

Método del paso constante	Método con paso $I/i$	Cantidad de iteraciones
0,0517934	0,0582595	100
0,0442324	0,0581383	200
0,0386819	0,0580719	300
0,0344183	0,0580300	400
0,0311954	0,0579990	500
0,0285636	0,0579747	600
0,0265812	0,0579548	700
0,0250686	0,0579374	800
0,0234637	0,0579229	900
0,0227099	0,0579084	1000
0,0218816	0,0578967	1000
0,0213104	0,0578850	1200
0,0205556	0,0578758	1300
0,0200188	0,0578657	1400
0,0197307	0,0578569	1500

Método del paso constante	Método con paso $l/i$	Cantidad de iteraciones
0,0192282	0,0578494	1600
0,0191537	0,0578411	1700
0,0188832	0,0578343	1800
0,0188705	0,0578289	1900
0,0188749	0,0578248	2000

En las tablas anteriores se ve que el método de paso constante desciende mucho más rápidamente, mientras que el otro desciende establemente, pero con mayor lentitud.

3. Los experimentos realizados dan lugar a las siguientes conclusiones:
  - (i) El algoritmo del gradiente con elección del paso según la regla de Armijo es el menos estable a los errores.
  - (ii) La efectividad de los métodos del paso constante y del del máximo descenso es aproximadamente la misma, pero en el segundo caso se logra en menor número de iteraciones.
  - (iii) En las etapas finales el método del mayor descenso funciona peor que el del paso constante.
  - (iv) El método con selección del paso  $l/i$  desciende establemente, pero con lentitud.
  - (v) Bajo perturbaciones aleatorias el campo de atracción resulta más estrecho que en el caso de perturbaciones indeterminadas del mismo nivel, pero al caer en un campo de estimación garantizado la velocidad de convergencia disminuye.
  
4. Como consecuencia de lo anterior se pueden hacer las siguientes recomendaciones:
  - (i) Es bueno comenzar utilizando la regla de Armijo (método 2) hasta que  $f(x_{n+1}) = f(x_n)$ , en donde  $n$  es algún número de iteración, es decir, mientras que el método 2 esté minimizando la función.
  - (ii) A continuación se puede alternar el método del máximo descenso con el del paso constante según el siguiente principio: del punto  $x$  en que nos encontremos hacemos una iteración con el método 3 y cinco iteraciones con el 1 (puesto que el método del mayor descenso es más caro que el del paso constante), y comparamos los valores de la función. Mediante el método con el cual el valor de

la función se hizo menor hacemos desde ese punto cien iteraciones si "ganó" el método 1, y veinte si "ganó" el 3; del nuevo punto obtenido avanzamos otra vez una iteración con el 3 y cinco con el 1, comparamos y así sucesivamente.

- (iii) En el método del paso constante se divide por 2 el paso cada 1.000 iteraciones.

Puesto que al elegir el paso según la regla de Armijo resulta, como se vio, una menor estabilidad a los errores, vamos a comparar los resultados de un experimento realizado según las recomendaciones anteriores con los obtenidos mediante la regla de Armijo. El experimento se realizó con la función de Rosenbroke

$$f(x) = 100 \sum_{i=2}^N (x_i - x_{i-1}^2) + (1 - x_1)^2,$$

con  $N = 10$  y  $N = 15$ , y con la función de Powell ya utilizada,

$$\begin{aligned} f: \mathbb{R}^4 &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) \equiv (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + \\ &+ 10(x_1 - x_4)^4. \end{aligned}$$

Como resultado del experimento se confirmaron las observaciones ya hechas. En el caso de la función de Powell las primeras 330 iteraciones coinciden, pero después de la iteración 350 comienza a prevalecer el método recomendado. En cuanto a las dos variantes con la función de Rosenbroke ( $N = 10$  y  $N = 15$ ), más o menos a partir de la iteración 27 el método 2 deja de actuar, mientras que el recomendado continúa permanentemente disminuyendo el valor de la función.

5. Se puede concluir que el método del gradiente funciona en general mejor sólo en las primeras etapas de búsqueda del mínimo, mientras los puntos  $x_n$  no estén demasiado cerca del punto optimal  $x^*$ . Esto está ligado al hecho de que en la vecindad del punto optimal el gradiente es muy pequeño, en particular su parte lineal (sobre la base de la cual se elige la dirección de descenso), así que se acrecienta la influencia de la parte cuadrática y el método se hace extremadamente sensible a las inevitables inexactitudes del cálculo. En consecuencia, cerca del punto minimal es necesario emplear métodos más exactos pero a la vez más complicados que tengan en cuenta los términos cuadráticos del incremento.

## Referencias

- [1] FIACCO A.V., MCCORMICK G.P. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley
- [2] MANGASARIAN OLVI L. *Nonlinear Programming*, SIAM, Philadelphia, 1994.
- [3] ARMIJO L. *Minimization of functions having continuous partial derivatives*. Pacific J. Math., 16 (1966), 1-3.
- [4] ZANGWILL WILLARD I. *Nonlinear Programming*. Prentice-Hall, Inc., Englewood Cliffs, W.J., 1969. ilustración 1 1