0.1 Modificación del método de Newton

El método de Newton, a partir de un punto inicial x^1, busca un punto crítico, es decir, un punto x^K tal que $f'(x^K) \approx 0$. Esto hace que para ciertas funciones este método no sea un método de minimización y entonces es posible que el punto crítico obtenido sea peor que el punto inicial, o sea,

$$
\begin{aligned}
&f'(x^K) \approx 0,\\
&f(x^K) > f(x^1).
\end{aligned}
$$

Por ejemplo, al tratar de minimizar

$$
\begin{aligned}
f(x_1, x_2) = &(x_1 + 1.5)(x_1 + 0.5(x_1 - 0.5) + (x_2 - 0.5)(x_2 - 1.5)(x_2 - 2.5)\\
&+ 0.3x_1x_2 + 0.01(x_1 - 3)^4 + 0.01(x_2 - 4)^4
\end{aligned}
$$

por el método de Newton a partir de $x^1 = (-2.5, 1)$, se obtiene:

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$		d^k
1	-2.5	3.5856	4.6450	-8.3700	0.3000	0.5190
	1.0		-2.0800	0.3000	-1.9200	-1.0022
2	-1.9810	5.0674	0.6360	-5.9086	0.3000	0.1266
	-0.0022		2.6115	0.3000	-7.0913	0.3736
3	-1.8544	5.6411	0.0386	-5.2984	0.3000	0.0112
	0.3714		0.3538	0.3000	-5.1916	0.0688
4	-1.8432	5.6538	0.0003	-5.2443	0.3000	0.0002
	0.4402		0.0122	0.3000	-4.8381	0.0025
5	-1.8430	5.6538	0.0000			
	0.4427		0.0000			

La modificación busca convertir el método en uno de descenso, es decir, siempre se debe tener que

$$f(x^{k+1}) < f(x^k).$$

En el método de Newton se obtiene d^k solución de $f''(x^k)\,d^k = -f'(x^k)$. Si $f(x^k + d^k) \geq f(x^k)$, entonces se averigua si d^k es dirección de descenso. Si lo es, se obtiene t_k minimizando $f(x^k + td^k)$ para $t \geq 0$ y se construye $x^{k+1} = x^k + t_k d^k$.

Si d^k no es dirección de descenso, en lugar de utilizar la matriz hessiana se utiliza

$$M = f''(x^k) + \lambda I, \tag{1}$$
$$Md^k = -f'(x^k). \tag{2}$$

Es claro que para $\lambda > 0$ suficientemente grande la matriz M es de diagonal positiva estrictamente dominante y, por lo tanto, definida positiva.

Recordemos que una matriz A simétrica es definida positiva si y solamente si su inversa existe y es definida positiva (los valores propios de A^{-1} son los inversos multiplicativos de los valores propios de A.)

La dirección d^k así obtenida es de descenso si

$$\begin{aligned} f'(x^k)^{\mathrm{T}} d^k &< 0, \\ f'(x^k)^{\mathrm{T}}(-M^{-1} f'(x^k)) &< 0, \\ -f'(x^k)^{\mathrm{T}} M^{-1} f'(x^k) &< 0. \end{aligned}$$

En consecuencia, si M es definida positiva, entonces d^k, obtenida según (2), es dirección de descenso.
Se podría pensar que la mejor estrategia es tomar inicialmente un valor λ grande para garantizar que M sea definida positiva. Esto hace que desde el principio M se aleje del hessiano, o sea, el método se aleja del método de Newton, perdiendo desde el comienzo las posibles ventajas del método de Newton, en particular la convergencia cuadrática.

Además, para valores grandes de λ, se tiene que $M \approx \lambda I$, o sea el método se aproximaría al método del descenso más pendiente y a su lentitud debida al zigzagueo.

Por eso se empieza con $M = f''(x^k)$ y, cuando no se tiene una dirección de descenso, se modifica M.

El siguiente valor de λ garantiza que $M = H + \bar{\lambda} I$ es definida positiva ($H = f''(x^k)$).

$$\sigma_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |h_{ij}|, \quad i = 1, ..., n$$

$$\sigma = \max_{1 \leq i \leq n} \{\sigma_i - h_{ii}\}$$

$$\lambda' > \max\{\sigma, 0\}$$

Si $\sigma < 0$, entonces H es definida positiva y no necesita modificaciones.

Método de Newton modificado

```
datos: x^1, ε_g, MAXIT
para k = 1, ...,MAXIT
    si ||f'(x^k)|| ≤ ε_g parar
    λ = 0
    fink = 0
    mientras fink = 0
        M = f''(x^k) + λI
        resolver Md^k = −f'(x^k)
        si f(x^k + d^k) < f(x^k)
            x^{k+1} = x^k + d^k
            fink = 1
        sino
            si f'(x^k)^T d^k < 0
                t_k = argmin f(x^k + td^k), t > 0
                x^{k+1} = x^k + t_k d^k
                fink = 1
            sino
                incrementar λ
            fin-si
        fin-si
    fin-mientras
fin-para
```

La terminación esperada del algoritmo anterior se tiene cuando $||f'(x^k)|| \leq \varepsilon_g$. Pero hay otras terminaciones posibles:

- demasiadas iteraciones

- $f(x^k + td^k)$ no tiene minimizador (no está acotada inferioremente).
- M no es invertible. En este caso se puede incrementar λ.

Ejemplo Minimizar

$$\begin{aligned} f(x_1, x_2) =&(x_1 + 1.5)(x_1 + 0.5(x_1 - 0.5) + (x_2 - 0.5)(x_2 - 1.5)(x_2 - 2.5) \\ &+ 0.3x_1x_2 + 0.01(x_1 - 3)^4 + 0.01(x_2 - 4)^4 \end{aligned}$$

por el método de Newton modificado a partir de $x^1 = (-2.5, 1)$.

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$		λ'
1	-2.5	3.5856	4.6450	-8.3700	0.3000	8.6700
	1.0		-2.0800	0.3000	-1.9200	

λ	M		d	$x^k + d$	$f(x^k + d)$	$f'(x^k)^{\mathrm{T}} d$	t_k
0	-8.37	0.30	0.5190	-1.9810	5.0674	4.4956	
	0.30	-1.92	1.0022	-0.0022			
2.89	-5.48	0.30	0.9490	-1.5510	3.9849	0.5581	
	0.30	0.97	1.8508	2.8508			
5.78	-2.59	0.30	1.8393	-0.6607	2.2385		
	0.30	3.86	0.3559	1.3959			

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$		λ'
2	-0.6607	2.2385	-2.4660	0.6439	0.3000	0.1108
	1.3959		-1.8721	0.3000	0.1892	

λ	M		d	$x^k + d$	$f(x^k + d)$	$f'(x^k)^{\mathrm{T}} d$	t_k
0	0.6439	0.3000	-2.9857	-3.6464	3232.0679	-20.0217	0.047556
	0.3000	0.1892	14.6278	16.0237			

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$		λ'
3	-0.8027	1.6105	-2.2972	-0.0809	0.3000	0.3809
	2.0915		-0.4691	0.3000	3.9863	

λ	M		d	$x^k + d$	$f(x^k + d)$	$f'(x^k)^{\mathrm{T}} d$	t_k
0	-0.0809	0.3000	-21.8620	-22.6647	-6543.7093		
	0.3000	3.9863	1.7629	3.8545			

Así se obtiene $x^4 = \begin{bmatrix} -22.6647 & 3.8545 \end{bmatrix}^{\mathrm{T}}$. Después de otras iteraciones

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$		λ'
6	-64.4160	-54550.7914	-0.0018	161.8936	0.3000	0
	4.1038		0.0140	0.3000	15.6239	

λ	M		d	$x^k + d$	$f(x^k + d)$	$f'(x^k)^{\mathrm{T}} d$	t_k
0	161.8936	0.3000	0.0000	-64.4159	-54550.7914		
	0.3000	15.6239	-0.0009	4.1029			

k	x^k	$f(x^k)$	$f'(x^k)$	$f''(x^k)$	λ'
7	-64.4159	-54550.7914	-0.0000		
	4.1029		0.0000		

0.2 Método de las tangentes paralelas

Este método es una modificación del método del descenso más pendiente. Sirve para evitar su zigzagueo. Dado un y^j, por medio del descenso más pendiente se obtiene z^j. En este punto z^j se utiliza la dirección $d = z^j - y^{j-1}$ para obtener y^{j+1}. Así sucesivamente hasta obtener y^{n+1}. Con este punto se vuelve a empezar. Obviamente, en la primera subiteración, y^2 se obtiene a partir de y^1 por medio del decenso más pendiente.

MÉTODO DE LAS TANGENTES PARALELAS

```
datos: x^1, ε_g, MAXIT
para k = 1,...,MAXIT
    y^1 = x^k
    si ||f'(y^1)|| ≤ ε_g parar
    d = -f'(y^1)
    t_1 = argmin f(y^1 + td), t ≥ 0
    y^2 = y^1 + t_1 d
    para j = 2,...,n
        si ||f'(y^j)|| ≤ ε_g parar
        d = -f'(y^j)
        t_j = argmin f(y^j + td), t ≥ 0
        z^j = y^j + t_j d
        d = z^j - y^{j-1}
        μ_j = argmin f(z^j + μd), μ ∈ ℝ
        y^{j+1} = z^j + μ_j d
    fin-para
    x^{k+1} = y^{n+1}
fin-para
```

Ejemplo Aplicar el método de tangentes paralelas para minimizar $f(x) = 0.1x_1^2 + x_2^2 + 10x_3^2 + 100x_4^2 - 0.2x_1 - 2x_2 - 20x_3 - 200x_4$, a partir del punto $x^1 = (2, 3, 4, 5)$.

j	y^j	$f'(y^j)$	d	t_j	z	d	μ_j
1	2	0.2	-0.2	0.005025			
	3	4	-4				
	4	60	-60				
	5	800	-800				
2	1.9990	0.1998	-0.1998	0.047812	1.9894	-0.0106	0.045515
	2.9799	3.9598	-3.9598		2.7906	-0.2094	
	3.6985	53.9695	-53.9695		1.1181	-2.8819	
	0.9797	-4.0676	4.0676		1.1741	-3.8259	
3	1.9890	0.1978	-0.1978	0.478224	1.8944	-0.1046	0.043371
	2.7810	3.5621	-3.5621		1.0776	-1.9023	
	0.9869	-0.2620	0.2620		1.1122	-2.5863	
	1.0000	0.0018	-0.0018		0.9992	0.0195	
4	1.8898	0.1780	-0.1780	4.865467	1.0239	-0.9650	0.024810
	0.9951	-0.0099	0.0099		1.0431	-1.7379	
	1.0000	0.0001	-0.0001		0.9997	0.0128	
	1.0000	0.0000	0.0000		1.0000	0.0000	
5	1.0000						
	1.0000						
	1.0000						
	1.0000						

Sigue una nueva iteración:

$$x^2 = y^5, \qquad y^1 = x^2, \qquad f'(y^1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Un resultado interesante y útil es el siguiente (ver la demostración en [Lue89]):

Para funciones cuadráticas estrictamente convexas, el método de tangentes paralelas es equivalente al método del gradiente conjugado. En particular, en ausencia de errores de redondeo y truncamiento, se obtiene la solución en no más de n subiteraciones.

Por ejemplo, al aplicar el método del gradiente conjugado a la función anterior se obtienen los siguientes resultados parciales

j	y^j	$f'(y^j)$	$°\alpha_j$	d^j	t_j
1	2	0.2		-0.2	0.005025
	3	4		-4	
	4	60		-60	
	5	800		-800	
2	1.9990	0.1998	0.004576	-0.2007	0.049988
	2.9799	3.9598		-3.9781	
	3.6985	53.9695		-54.2440	
	0.9797	-4.0676		0.4070	
3	1.9890	0.1978	0.004345	-0.1987	0.498965
	2.7810	3.5621		-3.5794	
	0.9869	-0.2620		0.0263	
	1.0000	0.0018		-0.0000	
4	1.8898	0.1780	0.002483	-0.1785	4.986180
	0.9951	-0.0099		0.0010	
	1.0000	0.0001		0.0000	
	1.0000	0.0000		0.0000	
5	1.0000				
	1.0000				
	1.0000				
	1.0000				

$$x^2 = y^5, \qquad y^1 = x^2, \qquad f'(y^1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$