

# Métodos de región confiable

Octavio Alberto Agustín Aquino

4 de junio de 2005

## Índice

<b>1. Bosquejo del algoritmo</b>	<b>2</b>
1.1. Aspectos generales . . . . .	2
1.2. Elección del radio . . . . .	2
<b>2. El punto de Cauchy y otros algoritmos</b>	<b>4</b>
2.1. Punto de Cauchy . . . . .	4
2.2. Método de la “pata de perro” . . . . .	5
<b>3. Código fuente</b>	<b>6</b>
3.1. Punto de Cauchy . . . . .	6
3.2. Pata de perro . . . . .	7
<b>4. Conclusiones</b>	<b>9</b>

## Índice de figuras

1. Gráfica de la corrida para <code>tabla=RegionConfiable(5,6,0.01,[1 -1]’,500)</code> . . . . .	6
2. Gráfica de la corrida para <code>tabla=pataperro(5,6,0.01,[1 -1]’,500)</code> . . . . .	7

# 1. Bosquejo del algoritmo

Los métodos de región confiable aproximan a la función objetivo con una función cuadrática en alguna región de radio apropiado, se halla el minimizador de la aproximación en tal región y se repite el proceso sobre el nuevo punto. Además, si la aproximación es exitosa, se expande el radio de la región, de lo contrario se reduce.

## 1.1. Aspectos generales

Sea  $f$  la función a minimizar. Asumiremos que los primeros dos términos del modelo cuadrático  $m_k$  en cada iteración  $x_k$  son idénticos a los primeros dos términos de la expansión de Taylor de  $f$  alrededor de  $x_k$ . Específicamente

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} \nabla p^T B_k p,$$

donde  $f_k = f(x_k)$ ,  $\nabla f_k = \nabla f(x_k)$ , y  $B_k$  es alguna matriz simétrica. Sabemos que

$$f(x_k + p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p.$$

para algún escalar  $f \in (0, 1)$ . La aproximación es de orden  $O(\|p\|^2)$ .

Para obtener el tamaño de paso, se busca una solución al subproblema

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \text{ sujeto a } \|p\| \leq \Delta_k \quad (1)$$

donde  $\Delta_k > 0$  es el radio de la región confiable. Cuando  $B_k$  es definida positiva y  $\|B_k^{-1} \nabla f_k\| \leq \Delta_k$ , la solución de (1) es simplemente  $p_k^B = -B_k^{-1} \nabla f_k$ . La solución en otros casos es menos obvia, pero usualmente puede hallarse sin mucho esfuerzo. En cualquier caso, basta con una solución aproximada para obtener la convergencia del método y un buen comportamiento en la práctica.

## 1.2. Elección del radio

Para elegir el radio de la región confiable  $\Delta_k$  en cada iteración nos basamos en la concordancia entre la función modelo  $m_k$  y la función objetivo

f en previas iteraciones. Dado el paso  $p_k$  definimos

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}. \quad (2)$$

Nótese que puesto que  $p_k$  se obtiene minimizando el modelo  $m_k$  sobre la región que incluye el paso  $p = 0$ , entonces el denominador de la fracción anterior siempre es no negativo. Por lo tanto, si  $\rho_k$  es negativo, el nuevo valor objetivo  $f(x_k + p_k)$  es mayor que el valor objetivo actual  $f(x_k)$ , así que el nuevo paso debe rechazarse.

Por otra parte, si  $\rho_k$  es cercano a 1, existe gran concordancia entre el modelo cuadrático y la función objetivo, así que es seguro expandir la región confiable en la siguiente iteración. Si  $\rho_k$  no es cercano a 1 no se altera la región confiable, pero si es cercano a 0 o negativo, contraemos la región confiable.

**Algoritmo 1.1 (Región confiable).** Esquema general del algoritmo de región confiable.

**Entrada:**  $\bar{\Delta}$ ,  $\Delta_0 \in (0, \bar{\Delta})$ ,  $\eta \in [0, \frac{1}{4})$  y  $\epsilon > 0$ .

- 1: **mientras**  $|x_{k+1} - x_k| > \epsilon$  **hacer**
- 2:   Obtener  $p_k$  resolviendo aproximadamente (1).
- 3:   Evaluar  $\rho_k$  de (2).
- 4:   **si**  $\rho_k < \frac{1}{4}$  **entonces**
- 5:      $\Delta_{k+1} = \frac{1}{4} \|p_k\|$ .
- 6:   **si no**
- 7:     **si**  $\rho_k > \frac{3}{4}$  y  $\|p_k\| = \Delta_k$  **entonces**
- 8:       $\Delta_k = \min(2\Delta_k, \bar{\Delta})$ .
- 9:     **si no**
- 10:       $\Delta_{k+1} = \Delta_k$
- 11:     **fin si**
- 12:   **fin si**
- 13:   **si**  $\rho_k > \eta$  **entonces**
- 14:      $x_{k+1} = x_k + p_k$ .
- 15:   **si no**
- 16:      $x_{k+1} = x_k$ .
- 17:   **fin si**

18: fin mientras

## 2. El punto de Cauchy y otros algoritmos

Para convertir el Algoritmo 1.1 en un algoritmo práctico, necesitamos resolver (1). No es necesario hallar el minimizador exacto en dicho problema, basta encontrar un punto en la región confiable que nos de la suficiente reducción.

### 2.1. Punto de Cauchy

La reducción suficiente puede cuantificarse en términos del punto de Cauchy, que denotaremos por  $p_k^C$  y que definiremos en términos del siguiente algoritmo.

**Algoritmo 2.1 (Cálculo del punto de Cauchy).** Halla una solución aproximada a (1).

- 1: Hallar el vector  $p_k^S$  que resuelve la versión lineal de (1):

$$p_k^S = \arg \min_{p \in \mathbb{R}^n} f_k + \nabla f_k^T p \text{ sujeto a } \|p\| \leq \delta_k. \quad (3)$$

- 2: Calcular el escalar  $\tau_k > 0$  que minimiza  $m_k(\tau p_k^S)$  dentro de la región confiable, esto es

$$\tau_k = \arg \min_{\tau > 0} m_k(\tau p_k^S) \text{ sujeto a } \|\tau p_k^S\| \leq \delta_k \quad (4)$$

- 3: Hacer  $p_k^C = \tau_k p_k^S$ .

De hecho, es fácil escribir una forma cerrada del punto de Cauchy. La solución a (3) es simplemente

$$p_k^S = -\frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k.$$

Finalmente,

$$p_k^C = -\tau_k \frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k,$$

donde

$$\tau_k = \begin{cases} 1 & \text{si } \nabla f_k^T B_k \nabla f_k \leq 0, \\ \text{mín}(\|\nabla f_k\|^3 \nabla f_k^T B_k \nabla f_k, 1) & \text{en otro caso.} \end{cases}$$

La ventaja del punto de Cauchy es que es fácil de calcular, pues no involucra operaciones matriciales, al mismo tiempo que proporciona una aproximación aceptable del tamaño de paso.

## 2.2. Método de la “pata de perro”

Cuando el radio de la región confiable  $\Delta$  es pequeño, la restricción  $\|p\| \leq \Delta$  asegura que el término cuadrático  $m$  tiene poco efecto en la solución de (1). En este caso, la verdadera solución  $p^B$  es aproximadamente la misma que la que obtendríamos de minimizar la función lineal  $f + g^T p$  sobre  $\|p\| \leq \Delta$ , esto es,

$$p^B \approx -\Delta \frac{g}{\|g\|}.$$

El método de la “pata de perro” encuentra una solución aproximada reemplazando lo que sería una trayectoria curva hacia el minimizador exacto de (1) por un camino que consiste en dos segmentos. El primer segmento corre del punto actual hacia el minimizador irrestricto a lo largo de la dirección de máximo descenso definida por

$$p^U = -\frac{g^T g}{g^T B g} g,$$

mientras que la segunda línea corre de  $p^U$  a  $p^B$ . Formalmente denotamos esta trayectoria por  $\tilde{p}(\tau)$  para  $\tau \in [0, 2]$ , donde

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U) & 1 \leq \tau \leq 2, \end{cases}$$

Sin embargo, el punto donde la trayectoria toca el borde de la región confiable puede calcularse exactamente, cuando existe tal punto. Si  $\|p^U\| \leq \Delta$ , entonces  $p = p^U$ , y la trayectoria no toca al borde. De lo contrario, resolvemos la ecuación cuadrática

$$\|p^U + (\tau - 1)(p^B - p^U)\|^2 = \Delta^2.$$

### 3. Código fuente

Todos los algoritmos se probaron con la función de Rosenbrock

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$

cuyo mínimo global está en  $(1, 1)$  y se programaron en Octave 2.1.50.

#### 3.1. Punto de Cauchy

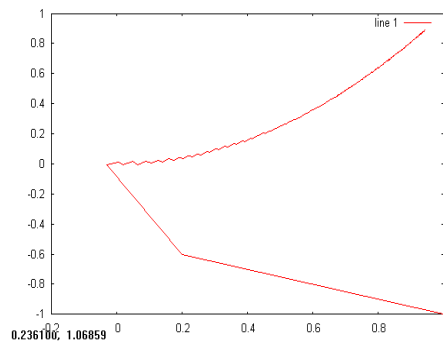


Figura 1: Gráfica de la corrida para `tabla=RegionConfiable(5,6,0.01,[1 -1]',500)`.

```
function tabla=RegionConfiable(delta,delta0,nu,x,maxit)
tabla=[x']
for i=1:maxit
hessiano=[-400*x(2)^2+1200*x(1)^2 -400*x(1); -400*x(1) 200];
gradiente=[-400*(x(2)-x(1)^2)*x(1)-2*(1-x(1)); 200*(x(2)-x(1)^2)];
aux=gradiente'*hessiano*gradiente;
if(aux>0)
tau=min((norm(gradiente))^3/(delta0*aux),1);
else
tau=1;
endif
endfor
```

```

p=(-tau*delta0)/norm(gradiente))*gradiente;
f0=100*(x(2)-x(1)^2)^2+(1-x(1))^2;
x=x+p;
f1=100*(x(2)-x(1)^2)^2+(1-x(1))^2;
m=-gradiente'*p + (1/2)*p'*hessiano*p;
rho=(f0-f1)/m;
if(rho<0.25)
    delta=norm(p)/4;
elseif ((rho>0.75) & (norm(p)==delta0))
    delta=min(2*delta0,delta)
endif
if(rho<=nu)
    x=x-p;
endif
tabla=[tabla; x'];
endfor
endfunction

```

### 3.2. Pata de perro

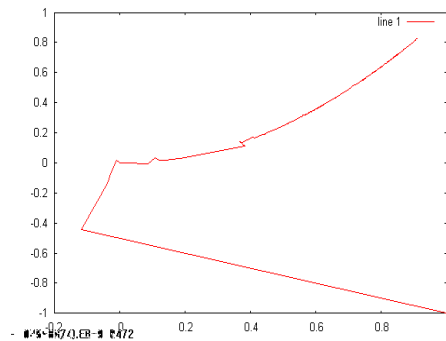


Figura 2: Gráfica de la corrida para `tabla=pataperro(5,6,0.01,[1 -1]',500)`.

```
function tabla=pataperro(delta,delta0,nu,x,maxit)
```

```

tabla=[x']
for i=1:maxit
    hessiano=[-400*x(2)^2+1200*x(1)^2 -400*x(1); -400*x(1) 200];
    gradiente=[-400*(x(2)-x(1)^2)*x(1)-2*(1-x(1)); 200*(x(2)-x(1)^2)];
    pU = -((gradiente'*gradiente)/(gradiente'*hessiano*gradiente))*gradiente;
    pB = -(delta/norm(gradiente))*gradiente;
    if(norm(pB) <= delta)
        p = pB;
    else
        a = pU'*pU + 2*pU'*pB + pB'*pB;
        b = -2*(pU'*pB+pB'*pB);
        c = (pB'*pB)-delta^2;
        tau = (-b+sqrt(b^2-4*a*c))/(2*a);
        p = pU +(tau-1)*(pB-pU);
    endif
    f0=100*(x(2)-x(1)^2)^2+(1-x(1))^2;
    x=x+p;
    f1=100*(x(2)-x(1)^2)^2+(1-x(1))^2;
    m=-gradiente'*p + (1/2)*p'*hessiano*p;
    rho=(f0-f1)/m;
    if(rho<0.25)
        delta=norm(p)/4;
    elseif ((rho>0.75) & (norm(p)==delta0))
        delta=min(2*delta0,delta)
    endif
    if(rho<=nu)
        x=x-p;
    endif
    tabla=[tabla; x'];
endfor
endfunction

```



## 4. Conclusiones

Vemos que el punto de Cauchy, en el caso de la función de Rosenbrock, fue ligeramente más efectivo que el de pata de perro, aunque la semejanza entre la manera en la que ambos convergen es notable.

Posiblemente los algoritmos de región confiable requieren mayor ajuste en el radio inicial y ciertas constantes involucradas en los algoritmos, pues puede verse que el comportamiento inicial de los algoritmos es muy pobre, y mejora mucho en las cercanías del óptimo.

## Referencias

- [1] Nocedal, Jorge *et al.* *Numerical optimization*, Springer-Verlag, Nueva York, 1999.