Optimización Numérica sin restricciones

Tema 4: Métodos de Región de Confianza

Oscar Dalmau

Centro de Investigación en Matemáticas
CIMAT

Marzo 2018

Orden del Tema

Orden del Tema

El Punto de Cauchy es el minimizador del modelo $m_k(p)$ a lo largo de de la dirección del máximo descendo, i.e., $p_k = -\lambda_k g_k$ sujeto a la región de confianza.

$$h(\lambda) \;:=\; m_k(-\lambda g_k) = f_k - g_k^T g_k \lambda + \frac{1}{2}\lambda^2 g_k^T B_k g_k;\ \lambda \geq 0$$

Como $\|p\| \leq \Delta_k$ entonces

$$\|-\lambda g_k\| \leq \Delta_k \;\Rightarrow \lambda \leq \frac{\Delta_k}{\|g_k\|} =: \bar{\lambda}$$

$$\boxed{\lambda_k = \arg \min_{\lambda \in [0,\bar{\lambda}]} h(\lambda)}$$

Cálculo del Punto de Cauchy

La solucion del problema anterior es

$$\begin{aligned}\lambda_k &= \begin{cases}\bar{\lambda}, & \text{si } g_k^T B_k g_k \leq 0 \\ \min\left(\bar{\lambda}, \frac{\|g_k\|^2}{g_k^T B_k g_k}\right), & \text{e.o.c.}\end{cases} \\ &= \bar{\lambda}\begin{cases}1, & \text{si } g_k^T B_k g_k \leq 0 \\ \min\left(1, \frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}\right), & \text{e.o.c.}\end{cases} \\ &= \bar{\lambda}\tau_k\end{aligned}$$

Resumiendo: $p_k^C = -\lambda_k g_k$. Luego $p_k^C = -\bar{\lambda}\tau_k g_k = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k$

Mejorando el Punto de Cauchy

- El Paso de Cauchy p_k^C produce un suficiente descenso en el modelo $m_k(\cdot)$ lo que permite obtener convergencia global.
- El costo computacional es bajo.
- Por las razones anteriores y teniendo en cuenta que usar el Paso de Cauchy p_k^C, es equivalente a usar el Método del Máximo descenso con un tamaño de paso particular, será posible mejorar la solución al problema cuadrático con restricciones original?

Mejorando el Punto de Cauchy

- El Paso de Cauchy p_k^C produce un suficiente descenso en el modelo $m_k(\cdot)$ lo que permite obtener convergencia global.
- El costo computacional es bajo.
- Por las razones anteriores y teniendo en cuenta que usar el Paso de Cauchy p_k^C, es equivalente a usar el Método del Máximo descenso con un tamaño de paso particular, será posible mejorar la solución al problema cuadrático con restricciones original?

Mejorando el Punto de Cauchy

- El Paso de Cauchy p_k^C produce un suficiente descenso en el modelo $m_k(\cdot)$ lo que permite obtener convergencia global.
- El costo computacional es bajo.
- Por las razones anteriores y teniendo en cuenta que usar el Paso de Cauchy p_k^C, es equivalente a usar el Método del Máximo descenso con un tamaño de paso particular, será posible mejorar la solución al problema cuadrático con restricciones original?

Mejorando el Punto de Cauchy

Observaciones

- El Paso de Cauchy p_k^C no depende fuertemente de la matriz B_k.
- Una mejor convergencia podría encontrarse si se usa la matriz B_k.
- Varios algoritmos de Región de Confianza primero calculan p_k^C y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

- El Paso de Cauchy p_k^C no depende fuertemente de la matriz B_k.
- Una mejor convergencia podría encontrarse si se usa la matriz B_k.
- Varios algoritmos de Región de Confianza primero calculan p_k^C y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

- El Paso de Cauchy p_k^C no depende fuertemente de la matriz B_k.
- Una mejor convergencia podría encontrarse si se usa la matriz B_k.
- Varios algoritmos de Región de Confianza primero calculan p_k^C y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

- Una estrategia simple es calcular el paso completo $p_k^B = -B_k^{-1}\nabla f_k$ siempre que B_k sea definida positiva
  - Si se cumple la restricción $\|p_k^B\| \leq \Delta_k$, entonces el tamaño del paso $p_k = p_k^B$
  - En otro caso se puede usar el paso de Cauchy $p_k = p_k^C$.
- Se puede hacer algo más?

Mejorando el Punto de Cauchy

Observaciones

- Una estrategia simple es calcular el paso completo $p_k^B = -B_k^{-1} \nabla f_k$ siempre que B_k sea definida positiva
  - Si se cumple la restricción $\|p_k^B\| \leq \Delta_k$, entonces el tamaño del paso $p_k = p_k^B$
  - En otro caso se puede usar el paso de Cauchy $p_k = p_k^C$.
- Se puede hacer algo más?

Mejorando el Punto de Cauchy

Observaciones

- Una estrategia simple es calcular el paso completo $p_k^B = -B_k^{-1}\nabla f_k$ siempre que B_k sea definida positiva
  - Si se cumple la restricción $\|p_k^B\| \leq \Delta_k$, entonces el tamaño del paso $p_k = p_k^B$
  - En otro caso se puede usar el paso de Cauchy $p_k = p_k^C$.
- Se puede hacer algo más? El Método Dogleg!

Método Dogleg

Idea del Algoritmo

- Se puede usar siempre y cuando la matriz B_k sea positiva definida, en otro caso usar el paso de Cauchy

Método Dogleg

Idea del Algoritmo

- Minimizar el model cuadrático sin restricciones a lo largo del gradiente: $p_k^U = \alpha \nabla f_k$
- Minimizar el model cuadrático sin restricciones si B_k es positiva definida: $p_k^B = -B_k^{-1} \nabla f_k$, i.e, obtener el paso completo
- Calcular el tamaño de paso $p_k = F(p_k^U, p_k^B)$, i.e., el tamaño de paso es una función que depende del *paso completo* y de la *dirección de máximo descenso*.

Trayectoria óptima $p(\Delta)$

Cual es el paso óptimo al variar la región de confianza?

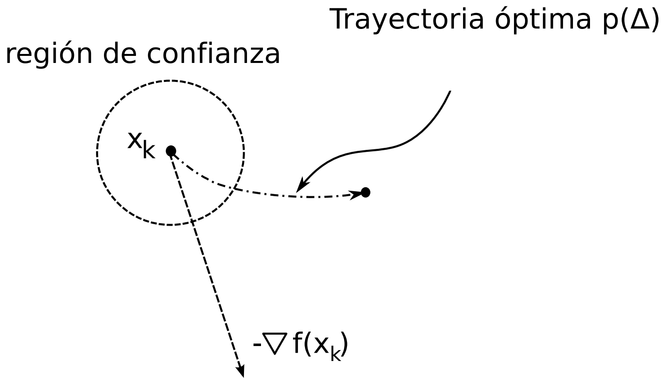

Figura: Trayectoria óptima

Trayectoria DogLeg

Aproximación de la trayectoria óptima usando el camino Dogleg

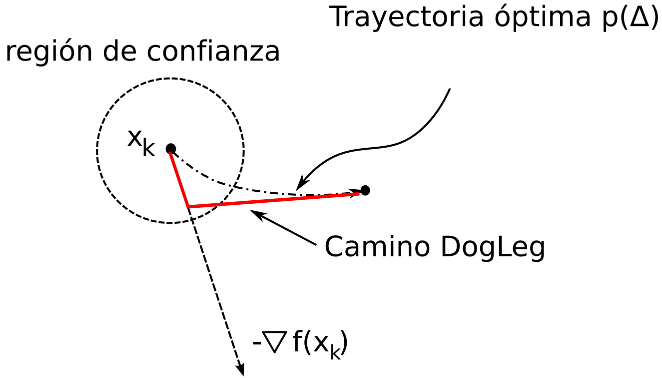

Figura: Camino o trayectoria Dogleg

Método Dogleg

Trayectoria Dogleg

- La primera línea del Dogleg va desde el origen hasta p_k^U
- La segunda línea va desde p_k^U hasta p_k^B, es decir, el Dogleg, $\tilde{p}(\tau)$, sigue la siguiente trayectoria

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

Método Dogleg

Minimizar el model cuadrático a lo largo del gradiente: p_k^U

- Si $p_k^U = \alpha \nabla f_k$, hallar el tamaño de paso α del problema sin restricciones (Unconstraint)

$$\begin{aligned} \alpha^* &= \arg\min_\alpha m_k(\alpha \nabla f_k) \\ m_k(\alpha \nabla f_k) &= f_k + \alpha \nabla f_k^T \nabla f_k + \frac{1}{2}\alpha^2 \nabla f_k^T B_k \nabla f_k \end{aligned}$$

$$\alpha^* = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T B_k \nabla f_k}$$

- Luego $p_k^U = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T B_k \nabla f_k} \nabla f_k$

Método Dogleg

Minimizar el model cuadrático a lo largo del gradiente: p_k^U

- Si $p_k^U = \alpha \nabla f_k$, hallar el tamaño de paso α del problema sin restricciones (Unconstraint)

$$\begin{aligned}\alpha^* &= \arg\min_\alpha m_k(\alpha \nabla f_k) \\ m_k(\alpha \nabla f_k) &= f_k + \alpha \nabla f_k^T \nabla f_k + \frac{1}{2}\alpha^2 \nabla f_k^T B_k \nabla f_k\end{aligned}$$

$$\alpha^* = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T B_k \nabla f_k}$$

- Luego $p_k^U = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T B_k \nabla f_k} \nabla f_k$

Método Dogleg

ww

Luego

- $p_k^U = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T B_k \nabla f_k} \nabla f_k$ y $p_k^B = -B_k^{-1} \nabla f_k$
- El problema consiste en encontrar el paso óptimo que minimiza el problema cuadrático con restricciones en la trayectoria Dogleg, es decir, en

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

- Como p_k^U y p_k^B son conocidos, entonces lo que tenemos que hallar es τ^*.

Método Dogleg

Lema

Si B es positiva definida. Entonces

- $\|\tilde{p}(\tau)\|$ es una función creciente de τ y
- $m(p(\tau))$ es una función decreciente de τ.

Método Dogleg

[i] El caso $\tau \in [0,1]$, entonces $p(\tau) = \tau p^U$, es trivial que $h(\tau) = \frac{1}{2}\|p(\tau)\|^2$ es creciente

Método Dogleg

[i] Consideremos el caso $\tau \in [1,2]$, cambiando variable $t = \tau - 1$, entonces $p(t) = p^U + t(p^B - P^U)$ con $t \in [0,1]$

$$\begin{aligned}
h(t) = \frac{1}{2}\|p(t)\|^2 &= \frac{1}{2}\|p^U\|^T + t(p^U)^T(p^B - P^U) + \frac{1}{2}t^2\|p^B - P^U\|^2 \\
h'(t) &= (p^U)^T(p^B - P^U) + t\|p^B - P^U\|^2 \\
&\geq (p^U)^T(p^B - P^U) \\
&= -\frac{g_k^T g_k}{g_k^T B g_k} g_k^T(-B^{-1}g_k + \frac{g_k^T g_k}{g_k^T B g_k} g_k) \\
&= \frac{g_k^T g_k}{g_k^T B g_k}(g_k^T B^{-1} g_k g_k^T B g_k - (g_k^T g_k)^2) \geq 0
\end{aligned}$$

pues $g_k^T B^{-1} g_k g_k^T B g_k \geq (g_k^T g_k)^2$, por Cauchy-Schwarz, luego $h'(t) \geq 0$ y por tanto $h(t)$ es creciente

Método Dogleg

[ii] Consideremos el caso $\tau \in [0, 1]$, entonces $p(\tau) = \tau p^U$

$$\begin{aligned}
m(p(\tau)) &= f_k + g_k^T p(\tau) + \frac{1}{2} p(\tau)^T B p(\tau) \\
m'(p(\tau)) &= (p'(\tau))^T g_k + (p'(\tau))^T B p(t) \\
&= (P^U)^T g_k + \tau (P^U)^T B P^U \\
&\leq (P^U)^T g_k + (P^U)^T B P^U \\
&\leq \frac{g_k^T g_k}{g_k^T B g_k} [-g_k^T g_k + \frac{g_k^T g_k}{g_k^T B g_k} g_k^T B g_k] = 0
\end{aligned}$$

luego $m'(p(t)) \leq 0$ y por tanto $m(p(t))$ es decreciente

Método Dogleg

[ii] Consideremos el caso $\tau \in [1, 2]$, cambiando variable $t = \tau - 1$, entonces $p(t) = p^U + t(p^B - P^U)$ con $t \in [0, 1]$

$$\begin{aligned}
m(p(t)) &= f_k + g_k^T p(t) + \frac{1}{2} p(t)^T B p(t) \\
m'(p(t)) &= (p'(t))^T g_k + (p'(t))^T B p(t) \\
&= (p^B - P^U)^T [g_k + BP^U + tB(p^B - P^U)] \\
&= (p^B - P^U)^T (g_k + BP^U) + t(p^B - P^U)^T B(p^B - P^U)] \\
&\leq (p^B - P^U)^T (g_k + BP^U) + (p^B - P^U)^T B(p^B - P^U) \\
&= (p^B - P^U)^T [g_k + BP^U + B(p^B - P^U)] \\
&= (p^B - P^U)^T [g_k + Bp^B] = 0
\end{aligned}$$

pues $g_k + Bp^B = 0$ luego $m'(p(t)) \leq 0$ y por tanto $m(p(t))$ es decreciente

Método Dogleg

Observaciones

Basados en el Lema anterior se tiene lo siguiente

- La trayectoria $\tilde{p}(\tau)$ intercepta a la *región de confianza*, $\|p\| = \Delta$, en un sólo punto si $\|p_k^B\| \geq \Delta$.
- Si $\|p_k^B\| \leq \Delta$ entonces el tamaño de paso óptimo es p_k^B, puesto que $m(p(\tau))$ decrece a lo largo del camino Dogleg. En otro caso hay que hallar el intercepto entre la trayectoria Dogleg y la region de confianza, por lo que se tiene que resolver la siguiente ecuación para τ

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

Método Dogleg

A partir de

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

Tenemos que resolver la ecuación cuadrática

$$a\lambda^2 + b\lambda + c \quad = \quad 0$$

donde $\tau = \lambda + 1$ y

$$a = \|p^B - p^U\|^2,\ b \quad = \quad 2(p^B)^T(p^B - p^U),\ c = \|p^U\|^2 - \Delta^2$$

Finalmente

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

Caso 1: DogLeg

p^B en en interior de la región de confianza

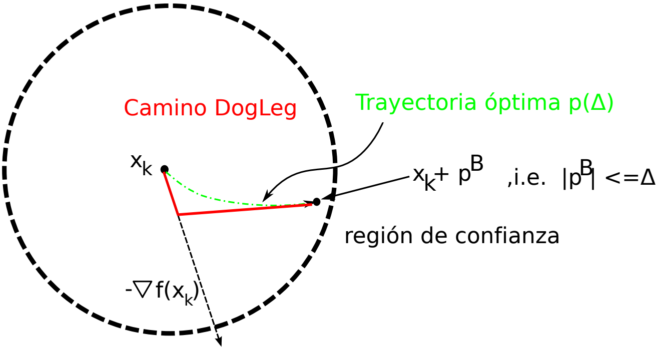

Figura:

Caso 2: DogLeg

p^B fuera de la región de confianza y $1 \leq \tau \leq 2$

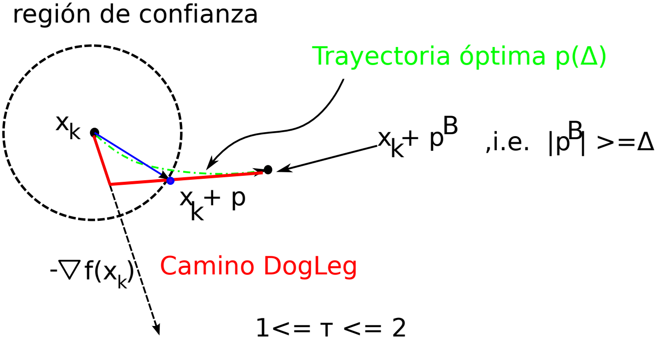

Figura:

Caso 3: DogLeg

p^B fuera de la región de confianza y $0 \leq \tau \leq 1$

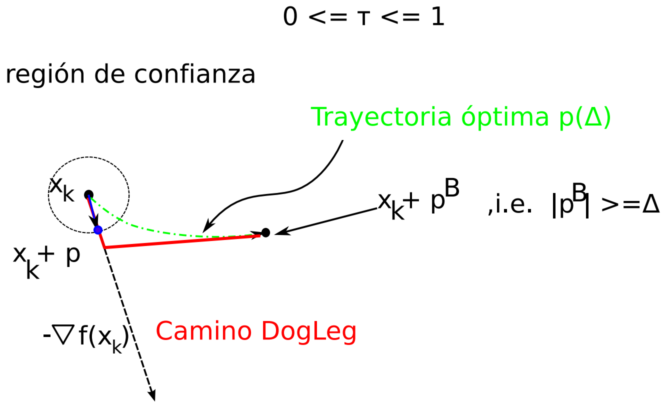

Figura:

Tarea

Implementar el Algoritmo de región de confianza usando las siguientes variantes para el cálculo del paso

- Usar el paso de Cauchy p^C.
- Si $\|p^B\| \leq \Delta$ usar p^B en caso contrario usar p^C.
- Si B es positiva definida usar el paso Dogleg en caso contrario usar p^C.

Aplicar cada una de las variantes anteriores para minimizar la función de Rosenbrock $f(x, y) = (1 - x)^2 + (y - x^2)^2$ usando puntos iniciales distintos. Realizar una tabla comparativa.