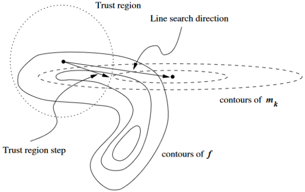
Métodos de Región de Confianza: Método Dog-Leg

Juan M. González-Campo
Pedro J. Marroquín

19 de Noviembre, 2024

Optimización sin Restricciones

En optimización sin restricciones, especialmente en problemas no lineales y de alta dimensión, la elección del tamaño de paso correcto es crucial. Pasos pequeños pueden hacer que la convergencia sea lenta, mientras que pasos grandes podrían alejarnos de las soluciones óptimas.



Región de confianza y búsqueda en línea

En cada iteración se define una región de confianza, como un vecindario alrededor de la solución actual. En esta región se espera que la función modelo refleje adecuadamente el comportamiento de la función objetivo.

Como función modelo generalmente se utilizá una función cuadrática:

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

Donde p es el paso tomado dentro de la región de confianza con radio Δ, f_k es el valor de la función objetivo en el punto actual x_k y B_k es una matriz que representa una aproximación al Hessiano (o el Hessiano exacto) de la función objetivo. El radio Δ se establece como un parámetro inicial y se ajusta en cada iteración en base a los resultados de la misma.

Entonces como subproblema se busca optimizar el modelo cuadrático, sujeto a la restricción $||p|| \leq \Delta$.

Ajuste del Radio

Como se mencionó anteriormente el radio Δ no es constante, se ajusta de acuerdo al resultado de las previas iteraciones. Para esto se utiliza una relación que explica el éxito que mide la reducción real de la función objetivo y se divide por la predicción de la reducción en el modelo.

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

Cuando la relación es cercana a 1 entonces se tiene un modelo preciso por lo que se puede expandir el radio permitiendo pasos más grandes y acelerando la convergencia, mientras que si difiere mucho de 1 se requiere más precisión por lo que se reduce el radio. En casos intermedios se suele dejar el radio sin ajustes, dependiendo de los umbrales de tolerancia que se definan.

Algoritmo General

Algoritmo 1.1 Región de Confianza (Actualización de Δ)

$\hat{\Delta} > 0, \Delta_0 \in (0, \hat{\Delta}), \eta \in [0, \frac{1}{4})$
while Criterio de paro **do**
  Obtener p_k minimizando (1) y calcular ρ_k con (2)
  if $\rho_k \leq \frac{1}{4}$ **then**
    $\Delta_{k+1} = \frac{1}{4}||p_k||$
  else
    if $\rho_k > \frac{3}{4}$ and $||p_k|| = \Delta_k$ **then**
      $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$
    else
      $\Delta_{k+1} = \Delta_k$

Algoritmo General (Continuación)

Algoritmo 1.2 Región de Confianza (Actualización de x)

while Criterio de paro (Continuación del while anterior) **do**
 if $\rho_k > \eta$ **then**
  $x_{k+1} = x_k + p_k$
 else
  $x_{k+1} = x_k$

$\hat{\Delta}$ es un valor arbitrario para definir un límite superior del radio de la región de confianza, Δ_0 es el radio inicial y η es un umbral arbitrario con el cual se considera que el paso p_k es suficientemente bueno para darse (si ρ_k es mayor a η).

Punto de Cauchy

El punto de Cauchy p_k^C minimiza m_k en la dirección de mayor decrecimiento $(-\nabla f_k)$; simplemente se limita a que esté dentro de la región de confianza $(||p_k^C|| \leq \Delta_k)$. p_k^C nos sirve como una aproximación de bajo costo y fácil de calcular de p_k que nos garantiza un reducción suficiente del modelo y por lo tanto lleva a convergencia global.

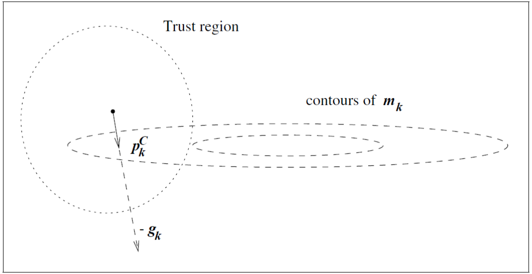


Figure: Representación visual del punto de Cauchy

Algoritmo para el punto de Cauchy

Algoritmo 2 Cálculo del punto de Cauchy (v.1)

Encontrar p_k^s que minimiza una versión lineal de (1):

$p_k^s = \arg\min_{p \in R^n} f_k + \nabla f_k^T p$ tal que $||p|| \leq \Delta_k$

Calcular $\tau_k > 0$ que minimice $m_k(\tau_k p_k^s)$ manteniéndose dentro de la región de confianza:

$\tau_k = \arg\min_{\tau > 0} m_k(\tau_k p_k^s)$ tal que $||\tau_k p_k^s|| \leq \Delta_k$

$p_k^C = \tau_k p_k^s$

Al resolver las ecuaciones dentro de este algoritmo se puede obtener una segunda versión para obtener el punto de Cauchy con operaciones definidas.

Algoritmo para el punto de Cauchy (Continuación)

Algoritmo 3 Cálculo del punto de Cauchy (v.2)

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

if $\nabla f_k^T B_k \nabla f_k \leq 0$ **then**

$\quad p_k^C = p_k^s$

else

$\quad \tau_k = \min\left(\frac{||\nabla f_k||^3}{\Delta_k \nabla f_k^T B_k \nabla f_k}, 1\right)$

$\quad p_k^C = \tau_k p_k^s$

return p_k^C

Punto de Cauchy Mejorado

Es importante notar que el punto de Cauchy p_k^C da una reducción suficiente a la función modelo que garantiza convergencia global. Sin embargo, simplemente utilizar el punto de Cauchy es una versión del método de descenso más empinado (Steepest Descent) con un tamaño de paso determinado, el cual se conoce que no tiene el mejor rendimiento incluso cuando se utiliza el punto óptimo de descenso.

Por esto, los métodos que buscan una solución p_k al subproblema de región de confianza utilizan el punto de Cauchy como un punto de partida, el cual mejoran para obtener convergencia acelerada. Para esto se puede reformular el subproblema de región de confianza como (sin los subíndices k para simplificar la notación):

$$\min_{p \in R^n} m(p) = f + g^T p + \frac{1}{2} p^T B p \tag{3}$$

donde $||p|| \leq \Delta$, con solución $p^*(\Delta)$ que enfatiza la dependencia de Δ y $g^T = \nabla f(x_k)$.

El Método Dogleg

El método Dogleg es uno de los que utilizan el punto de Cauchy e intentan mejorarlo. Para esto se evalúa el efecto del radio de la región de confianza Δ en la solución $p^*(\Delta)$, del subproblema (3).

Cuando B es positivo definida, entonces el minimizador de m es el paso completo $p^B = -B^{-1}g$. Cuando este es factible en (3), es una solución, por lo que $p^*(\Delta) = p^B$ siempre y cuando $\Delta \geq ||p^B||$.

Además cuando Δ es pequeño tenemos que $p^*(\Delta) \approx -\Delta \frac{g}{||g||}$.

Para valores intermedios de Δ la solución $p^*(\Delta)$ sigue una trayectoria curvada.

El Método Dogleg

El método Dogleg reemplaza la trayectoria curvada por dos segmentos de línea. El primero sigue desde el origen hasta el minimizador de descenso más empinado definido por $p^U = -\frac{g^T g}{g^T B g} g$.

El segundo segmento comienza en p^U y termina en p^B. Formalmente se denota la trayectoria $\tilde{p}(\tau)$ para $\tau \in [0, 2]$, donde

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

El método escoge p para minimizar el modelo m en esta trayectoria, sujeto a la acotación de la región de confianza.

Trayectoria del método Dogleg

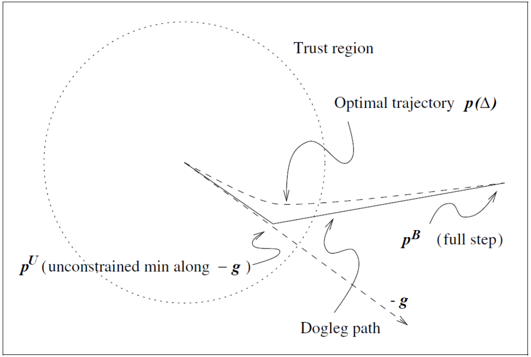


Figure: Trayectoria del método Dogleg

Solución del Método Dogleg

Es de notar que no es necesario realizar una busqueda en el método, ya que la trayectoria Dogleg intersecta a lo más 1 vez la región de confianza y es posible computar esta intersección analíticamente.

Lemma

Sea B positiva definida. Entonces:
(i) $||\tilde{p}(\tau)||$ *es una función creciente de* τ *y*
(ii) $m(\tilde{p}(\tau))$ *es una función decreciente de* τ*.*

Dado esto, la trayectoria $\tilde{p}(\tau)$ intersecta con el límite de la región de confianza $||p|| = \Delta$ exactamente en un punto si $||p^B|| \geq \Delta$ y en ninguno en cualquier otro caso.

Solución del Método Dogleg

Como m es decreciente en la trayectoria, el valor de p será p^B si $||p^B|| \leq \Delta$, en caso contrario será en el punto de intersección de la trayectoria y el límite de la región de confianza. En este caso se calcula el valor apropiado de τ resolviendo

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

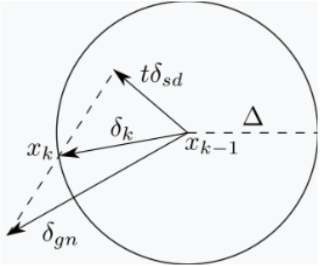


Figure: Paso de Dogleg

Subalgoritmo para Dogleg

Ya que el método de Dogleg se enfoca en calcular un p_k, el algoritmo completo del método simplemente debe ser una adaptación del Algoritmo General descrito anteriormente.

Algoritmo 4 Subalgoritmo Dogleg

$p^U = -\frac{g^T g}{g^T B g} g$
$p^B = -B^{-1} g$
if $||p^B|| \leq \Delta$ **then**
  return p^B
else if $||p^U|| \geq \Delta$ and $||p^B|| \geq \Delta$ **then**
  return $-\Delta \frac{p^U}{||p^U||}$
else
  return $p^U + (\tau - 1)(p^B - p^U)$ tal que intersecta la región de confianza

Ejemplo: Minimizar la función de Rossenbrock

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \quad (4)$$

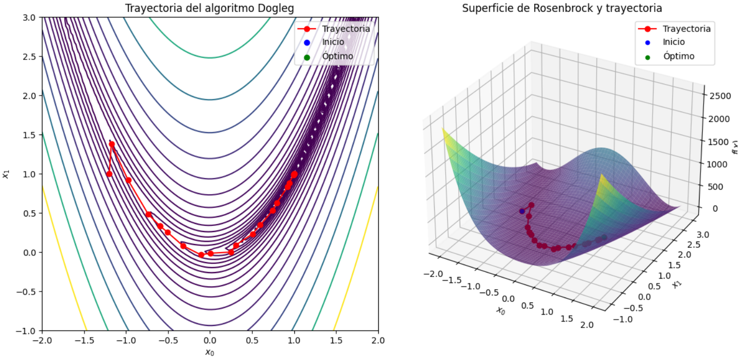


Figure: Visualización de la solución del método Dog-Leg

Comparación: Descenso Gradiente

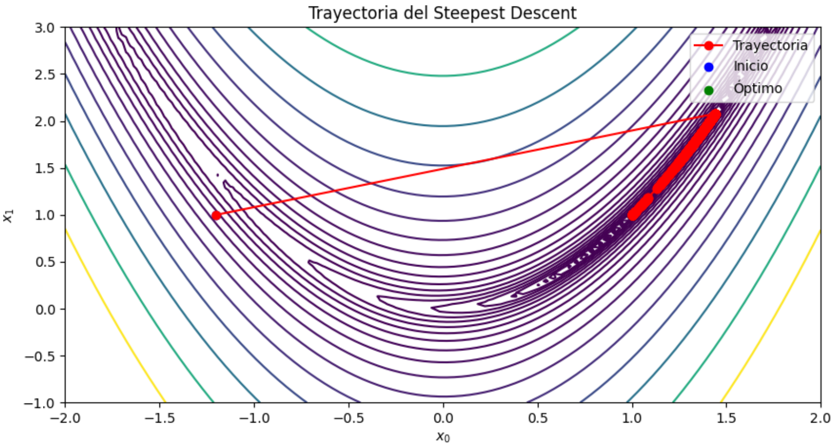


Figure: Trayectoria Steepest Descent con Golden Ratio