Optimización Numérica sin restricciones

Tema 4: Métodos de Región de Confianza

Oscar Dalmau

Centro de Investigación en Matemáticas
CIMAT

24 de marzo de 2021

Tema 4

Métodos de Región de Confianza

Orden del Tema

Orden del Tema

1. Métodos de Región de Confianza
   - Introducción: Idea General
   - Punto de Cauchy

Características generales

- Los métodos de región de confianza (MRC), similar a los métodos de búsqueda en línea, para *generar los pasos* se basan en un *modelo que aproxima la función objetivo*.
- Los MRC, en cada iteración, definen una región en la cual se confía que el modelo se ajusta bien a la función, a esta región se le denomina *región de confianza*.

Características generales

- Los métodos de región de confianza (MRC), similar a los métodos de búsqueda en línea, para *generar los pasos* se basan en un *modelo que aproxima la función objetivo*.
- Los MRC, en cada iteración, definen una región en la cual se confía que el modelo se ajusta bien a la función, a esta región se le denomina *región de confianza*.

Representación gráfica

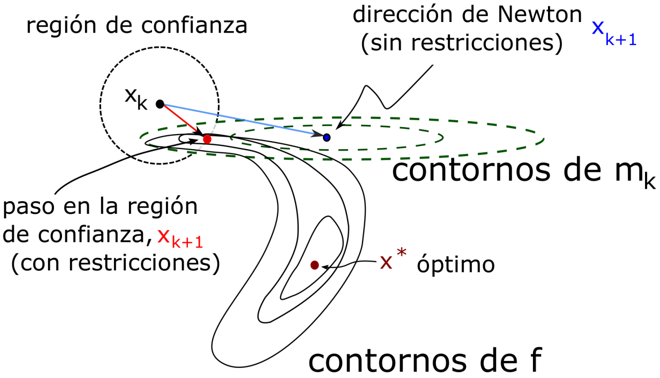

Figura: Comparación de pasos calculados usando la región de confianza y mediante el Método de Newton.

Características generales

- *El paso se calcula como un minimizador aproximado del modelo, $m_k(p)$, restringido a la región de confianza ($\mathcal{R}_c$).*

$$p_k = \arg\min_p m_k(p), \text{ s.t. } p \in \mathcal{R}_c.$$

- *La dirección de descenso y el tamaño de paso se calculan al mismo tiempo, diferente a los métodos de búsqueda en línea que primero se buscan una dirección de descenso y luego el tamaño de paso.*

$$x_{k+1} = x_k + p_k$$

Características generales

- *El paso se calcula como un minimizador aproximado del modelo, $m_k(p)$, restringido a la región de confianza ($\mathcal{R}_c$).*

$$p_k = \arg\min_p m_k(p), \text{ s.t. } p \in \mathcal{R}_c.$$

- *La dirección de descenso y el tamaño de paso se calculan al mismo tiempo, diferente a los métodos de búsqueda en línea que primero se buscan una dirección de descenso y luego el tamaño de paso.*

$$x_{k+1} = x_k + p_k$$

Características generales

- Si el paso obtenido usando el método de región de confianza no produce un progreso en la minimización entonces el paso no es aceptado.
- Lo anterior es un indicativo de que *el modelo no se ajustó bien a la función en la región de confianza y entonces se reduce el tamaño de la region de confianza y se recalcula el paso usando la nueva región de confianza.*

Características generales

- Si el paso obtenido usando el método de región de confianza no produce un progreso en la minimización entonces el paso no es aceptado.
- Lo anterior es un indicativo de que *el modelo no se ajustó bien a la función en la región de confianza y entonces se reduce el tamaño de la region de confianza y se recalcula el paso usando la nueva región de confianza.*

Características generales

Observaciones

- Si la región de confianza es muy pequeña, entonces el tamaño de paso será muy pequeño, y por tanto, podríamos acercarnos lentamente al óptimo local y el costo computacional (número de iteraciones) podría aumentar considerablemente.
- Si la región de confianza es muy grande entonces el paso podría conducir a un punto que este muy alejado del óptimo de la función. Por lo tanto, habría que reducir la región de confianza muchas veces antes de obtener un tamaño de paso adecuado. Y esto atentaría también en contra el costo computacional.

Características generales

Observaciones

- Si la región de confianza es muy pequeña, entonces el tamaño de paso será muy pequeño, y por tanto, podríamos acercarnos lentamente al óptimo local y el costo computacional (número de iteraciones) podría aumentar considerablemente.
- Si la región de confianza es muy grande entonces el paso podría conducir a un punto que este muy alejado del óptimo de la función. Por lo tanto, habría que reducir la región de confianza muchas veces antes de obtener un tamaño de paso adecuado. Y esto atentaría también en contra el costo computacional.

Qué necesitamos?

Ingredientes para los Algoritmos de región de confianza

- Un *modelo*.
- *Radio* o *tamaño* de la región de confianza.
- *Una medida para evaluar el ajuste del modelo en la región de confianza.*

Qué necesitamos?

Ingredientes para los Algoritmos de región de confianza

- Un *modelo*.
- *Radio* *o* *tamaño* *de la región de confianza.*
- *Una medida para evaluar el ajuste del modelo en la región de confianza.*

Qué necesitamos?

Ingredientes para los Algoritmos de región de confianza

- Un *modelo*.
- *Radio* *o* *tamaño* *de la región de confianza.*
- *Una medida para evaluar el ajuste del modelo en la región de confianza.*

Qué modelo?

- Se asumirá que el modelo m_k que será usado en la iteración x_k es cuadrático.
- Usando la aproximación de Taylor de segundo orden, es decir,

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

con $t \in (0, 1)$. Y tomando una aproximaxión del Hessiano, se tiene que

$$m_k(p) \quad = \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

Qué modelo?

- Si B_k es una matriz simétrica, entonces la diferencia entre la función y el modelo es del orden $O(\|p\|^2)$.
- Si B_k es el Hessiano $\nabla^2 f(x_k)$, entonces la diferencia entre la función y el modelo es del orden $O(\|p\|^3)$.

Modelo para la obtención del paso

Obtención del paso usando MRC

- Para hallar el paso, se resuelve el siguiente problema de opimimización con restricciones:

$$\begin{aligned} p_k^* = \arg\min\nolimits_p m_k(p) \quad &= \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\ \text{s.t.} \quad & \|p\| \leq \Delta_k, \end{aligned}$$

donde Δ_k es el *radio de la región de confianza*.

Modelo para la obtención del paso

Obtención del paso usando MRC

- Si B_k es positiva definida, la solución del problema sin restricciones

$$\arg\min_p m_k(p) \quad = \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

  se obtiene directamente mediante $p^* = B_k^{-1} \nabla f(x_k)$.
- Si además se cumple que $\|p^*\| \leq \Delta_k$, i.e., $\|B_k^{-1} \nabla f(x_k)\| \leq \Delta_k$, entonces la solución del problema de optimización con restricciones original es el *paso completo*, es decir $p_k^* = B_k^{-1} \nabla f(x_k)$.

Modelo para la obtención del paso

Obtención del paso usando MRC

- En cualquier otro caso, la solución del problema

$$\begin{aligned} p_k^* = \arg\min_p m_k(p) &= f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\ \text{s.t.} & \quad \|p\| \leq \Delta_k, \end{aligned}$$

  puede ser muy difícil.

- En la práctica no se necesita resolver completamente el problema anterior y, por lo general, una solución aproximada al problema anterior es suficiente.

Modelo para la obtención del paso

Obtención del paso usando MRC

- En cualquier otro caso, la solución del problema

$$\begin{aligned} p_k^* = \arg \min\nolimits_p m_k(p) \quad &= \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\ \text{s.t.} \quad & \|p\| \leq \Delta_k, \end{aligned}$$

  puede ser muy difícil.
- En la práctica no se necesita resolver completamente el problema anterior y, por lo general, una solución aproximada al problema anterior es suficiente.

Como evaluar la calidad del modelo?

Calidad del modelo y radio de la región de confianza

- Un elemento importante en los métodos de región verdadera es la forma de calcular el radio de la region de confianza Δ_k en cada iteración.
- Para ello se calcula la siguiente medida del ajuste

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

Donde el numerador $f(x_k) - f(x_k + p_k)$ representa la reducción en la función, y el denominador $m_k(0) - m_k(p_k)$ la reducción en el modelo.

Como evaluar la calidad del modelo?

Sobre la medida

- Medida

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

- La reducción en el modelo, i.e., el denominador $m_k(0) - m_k(p_k)$, siempre positivo, pues p_k minimiza el modelo.
- Si $\rho_k < 0$ entonces $f(x_k) < f(x_k + p_k)$ y la función se incrementa en lugar de decrementarse. En este caso, en el algoritmo se debe rechazar el paso p_k.

Como evaluar la calidad del modelo?

Sobre la medida

- Medida

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

- La reducción en el modelo, i.e., el denominador $m_k(0) - m_k(p_k)$, siempre positivo, pues p_k minimiza el modelo.
- Si $\rho_k < 0$ entonces $f(x_k) < f(x_k + p_k)$ y la función se incrementa en lugar de decrementarse. En este caso, en el algoritmo se debe rechazar el paso p_k.

Como evaluar la calidad del modelo?

Sobre la medida

- Medida

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

- La reducción en el modelo, i.e., el denominador $m_k(0) - m_k(p_k)$, siempre positivo, pues p_k minimiza el modelo.
- Si $\rho_k < 0$ entonces $f(x_k) < f(x_k + p_k)$ y la función se incrementa en lugar de decrementarse. En este caso, en el algoritmo se debe rechazar el paso p_k.

Como evaluar la calidad del modelo?

Sobre el radio de la región de confianza

- Medida

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

- Si $\rho_k \approx 1$ entonces el comportamiento de la función y el modelo concuerdan bastante bien en esta iteración, y es buena idea incrementar el radio de la región de confianza en la próxima iteración.
- Si ρ_k es positivo menor que 1, *pero cercano a 1, entonces no se modifica el radio Δ_k en la próxima iteración, i.e.,* $\Delta_{k+1} = \Delta_k$.

Como evaluar la calidad del modelo?

Sobre el radio de la región de confianza

- Medida

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

- Si $\rho_k \approx 1$ entonces el comportamiento de la función y el modelo concuerdan bastante bien en esta iteración, y es buena idea incrementar el radio de la región de confianza en la próxima iteración.
- Si ρ_k es positivo menor que 1, *pero cercano a 1, entonces no se modifica el radio Δ_k en la próxima iteración, i.e.,* $\Delta_{k+1} = \Delta_k$.

Como evaluar la calidad del modelo?

Sobre el radio de la región de confianza

- Medida

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

- Si ρ_k es positivo cercano a cero o negativo entonces se reduce el Δ_k en la próxima iteración, puesto que la función incrementó su valor o el ajuste del modelo no es bueno.

Algoritmo Región de confianza

- Dado $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \frac{1}{4})$
- Para $k = 0, 1, 2, \cdots$,
  - **Calcular** p_k (minimizar el modelo cuadático en la region de confianza, ie, $\|p\| \leq \Delta_k$)
  - **Calcular** ρ_k .
  - Si $\rho_k > \eta$ entonces $x_{k+1} = x_k + p_k$
    De lo contrario $x_{k+1} = x_k$
  - **Calcular** Δ_{k+1}, dados $(p_k, \Delta_k, \hat{\Delta}, \eta)$.

Algoritmo

- Dado $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \frac{1}{4})$
- Para $k = 0, 1, 2, \cdots,$
  - Se obtiene una solución aproximada del problema cuadrático con restricciones (se verá en próximas clases).
  - Se calcula ρ_k
  - Si $\rho_k < \frac{1}{4}$ entonces $\Delta_{k+1} = \frac{1}{4}\Delta_k$.
    De lo contrario, (i.e. $\rho_k \geq \frac{1}{4}$)
    - Si $\rho_k > \frac{3}{4}$ y $\|p_k\| = \Delta_k$ entonces $\Delta_{k+1} = min\{2\Delta_k, \hat{\Delta}\}$.
      De lo contrario $\Delta_{k+1} = \Delta_k$ (i.e. $\rho_k \in [\frac{1}{4}, \frac{3}{4}]$ o $\|p_k\| < \Delta_k$)

    Fin Si
  - Si $\rho_k > \eta$ entonces $x_{k+1} = x_k + p_k$
    De lo contrario $x_{k+1} = x_k$

Algoritmo

- Dado $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \eta_1]$
- Para $k = 0, 1, 2, \cdots$,
  - Se obtiene una solución aproximada del problema cuadrático con restricciones (se verá en próximas clases).
  - Se calcula ρ_k
  - Si $\rho_k < \eta_1$ entonces $\Delta_{k+1} = \hat{\eta}_1 \Delta_k$.
    De lo contrario, (i.e. $\rho_k \geq \eta_1$)
    - Si $\rho_k > \eta_2$ y $\|p_k\| = \Delta_k$ entonces $\Delta_{k+1} = min\{\hat{\eta}_2 \Delta_k, \hat{\Delta}\}$.
      De lo contrario $\Delta_{k+1} = \Delta_k$ (i.e. $\rho_k \in [\eta_1, \eta_2]$ o $\|p_k\| < \Delta_k$)

    Fin Si
  - Si $\rho_k > \eta$ entonces $x_{k+1} = x_k + p_k$
    De lo contrario $x_{k+1} = x_k$

Valores por defecto: $\eta_1 = \frac{1}{4}$, $\eta_2 = \frac{3}{4}$, $\hat{\eta}_1 = \frac{1}{4}$, $\hat{\eta}_2 = 2$

Algoritmo

Comentarios sobre el Algoritmo

- Se incrementa el radio de la región de confianza solamente, si concuerda el modelo con la función, i.e., $\rho_k > \frac{3}{4}$ y si al mismo tiempo p_k alcanza el borde de la región de confianza, i.e., $\|p_k\| = \Delta_k$
- Si el paso está en el interior de la región de confianza, entonces se concluye que el radio de la región de confianza no interfiere con el progreso del algoritmo y por tanto se deja el radio sin modificar.
- El radio solo se reduce, si la función incrementa su valor, o si el modelo y la función no concuerdan bien, i.e., $\rho_k < \frac{1}{4}$.

¿Qué falta para completar el Algoritmo?

¿Cómo calcular del Paso?

- Para hallar el paso, se resuelve el siguiente problema de opimimización con restricciones:

$$p_k^* = \arg\,\mathsf{min}_p m_k(p) \quad = \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p,$$
$$\text{s.t.} \quad \|p\| \leq \Delta_k.$$

Donde Δ_k es el radio de la región de confianza.

- Como se comentó anteriormente, la solución del problema anterior puede ser muy compleja.

¿Qué falta para completar el Algoritmo?

¿Cómo calcular del Paso?

- Para hallar el paso, se resuelve el siguiente problema de opimimización con restricciones:

$$\begin{aligned} p_k^* = \arg\,\mathsf{min}_p m_k(p) &= f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\ \text{s.t.} &\quad \|p\| \leq \Delta_k. \end{aligned}$$

  Donde Δ_k es el radio de la región de confianza.

- Como se comentó anteriormente, la solución del problema anterior puede ser muy compleja.

Sobre el calculo del Paso

Teorema

El vector p^* es una solución global del problema

$$\min_p m_k(p) \quad = \quad f(x_k) + g^T p + \frac{1}{2} p^T B p, \text{ s.t. } \|p\| \leq \Delta.$$

si y solo si, p^* es factible y existe $\lambda \geq 0$ y ademas se cumplen las condiciones

$$\begin{aligned} (B + \lambda I)p^* &= -g \\ \lambda(\|p^*\| - \Delta) &= 0 \\ B + \lambda I &\succeq 0 \end{aligned}$$

Nota: Ver detalles en el curso optimización II

Sobre el calculo del Paso: Comentarios

- Si $\lambda = 0$ entonces $Bp^* = -g$ (en particular si $B \succ 0$ entonces $p^* = -B^{-1}g$, ie, es el paso de Newton o Newton aproximado).
- Si $\lambda > 0$ entonces $Bp^* + \lambda p^* = -g$ por lo que $\lambda p^* = -(Bp^* + g)$, es decir, $p^* = -\frac{1}{\lambda}\nabla m_k(p^*)$ y por tanto p^* y $\nabla m_k(p^*)$ son paralelos, ie, p^* es la direccion de maximo descenso del modelo.
  Por otro lado, por la condición, $\lambda(\|p^*\| - \Delta) = 0$ se tiene que cumplir $\|p^*\| = \Delta$, es decir, la solución esta en la frontera (se activa la restricción)
- Vamos a describir una solución que aproxima el subproblems anterior, el cual obtiene una reducción del modelo m_k en al menos la que se obtiene a través del punto de Cauchy.

Sobre el calculo del Paso

- Una alternativa para aproximar la solución del problema anterior se basa en el punto de Cauchy.
- Una estrategia de aproximación es el método dogleg, el cual es una aproximación cuando B_k es definida positiva.

Orden del Tema

Punto de Cauchy

Definición

El Punto de Cauchy es el minimizador del modelo m_k a lo largo de la dirección del máximo descenso de la funcion f, i.e., $-\nabla f(x_k)$, sujeto a la región de confianza.

Alternativa para hallar el paso

La alternativa de solución del problema de optimización para hallar el paso recibe el nombre del Método Dogleg (Próxima clase) y está basada en el cáculo de el Punto de Cauchy .

Punto de Cauchy

- Para hallar el paso, se resuelve el problema de opimimización con restricciones:

$$p_k^* = \arg\min_p m_k(p) \quad = \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p,$$
$$\text{s.t.} \quad \|p\| \leq \Delta_k.$$

  Donde Δ_k es el radio de la región de confianza.

- Aunque en principio uno busca la solución del problema anterior, en la práctica, es suficiente encontrar una aproximación de p_k en la región de confianza que de un suficiente descenso del modelo para garantizar una convergencia del método.

Punto de Cauchy

- Para hallar el paso, se resuelve el problema de opimimización con restricciones:

$$\begin{aligned} p_k^* = \arg\min{}_p m_k(p) \quad &= \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\ \text{s.t.} \quad & \|p\| \leq \Delta_k. \end{aligned}$$

  Donde Δ_k es el radio de la región de confianza.

- Aunque en principio uno busca la solución del problema anterior, en la práctica, es suficiente encontrar una aproximación de p_k en la región de confianza que de un suficiente descenso del modelo para garantizar una convergencia del método.

Punto de Cauchy

- Para hallar el paso, se resuelve el problema de opimimización con restricciones:

$$
\begin{aligned}
p_k^* = \arg\min_p m_k(p) \quad &= \quad f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \\
\text{s.t.} \quad &\|p\| \leq \Delta_k.
\end{aligned}
$$

  Donde Δ_k es el radio de la región de confianza.
- El Punto de Cauchy, denotado como p_k^C, nos permite cuantificar el suficiente descenso del modelo.

Cálculo del Punto de Cauchy

Algoritmo: Punto de Cauchy

- Encontrar el punto p_k^S que resuelve la versión lineal:

$$p_k^S = \arg\min_p f(x_k) + \nabla f(x_k)^T p, \ \text{s.t.} \ \|p\| \leq \Delta_k$$

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \ \text{s.t.} \ \|\tau p_k^S\| \leq \Delta_k$$

- Calcular el Punto de Cauchy haciendo $p_k^C = \tau_k p_k^S$.

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 1)

- Encontrar el punto p_k^S que resuelve la versión lineal:

$$p_k^S = \arg\min_p f(x_k) + \nabla f(x_k)^T p, \text{ s.t. } \|p\| \leq \Delta_k$$

- La función decrece a lo largo de $-\nabla f(x_k)^T$, luego $p_k^S = -\lambda \nabla f(x_k)$ con $\lambda > 0$
- Como $\|p_k^S\| \leq \Delta_k$, entonces $\lambda \leq \frac{\Delta_k}{\|\nabla f(x_k)\|}$
- El máximo descenso se obtiene para $\lambda = \frac{\Delta_k}{\|\nabla f(x_k)\|}$, por lo que $p_k^S = -\frac{\Delta_k}{\|\nabla f(x_k)\|}\nabla f(x_k)$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 1)

- Encontrar el punto p_k^S que resuelve la versión lineal:

$$p_k^S = \arg\min_p f(x_k) + \nabla f(x_k)^T p, \text{ s.t. } \|p\| \leq \Delta_k$$

- La función decrece a lo largo de $-\nabla f(x_k)^T$, luego $p_k^S = -\lambda \nabla f(x_k)$ con $\lambda > 0$
- Como $\|p_k^S\| \leq \Delta_k$, entonces $\lambda \leq \frac{\Delta_k}{\|\nabla f(x_k)\|}$
- El máximo descenso se obtiene para $\lambda = \frac{\Delta_k}{\|\nabla f(x_k)\|}$, por lo que $p_k^S = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 1)

- Encontrar el punto p_k^S que resuelve la versión lineal:

$$p_k^S = \arg\min_p f(x_k) + \nabla f(x_k)^T p, \ \text{s.t.} \ \|p\| \leq \Delta_k$$

- La función decrece a lo largo de $-\nabla f(x_k)^T$, luego $p_k^S = -\lambda \nabla f(x_k)$ con $\lambda > 0$
- Como $\|p_k^S\| \leq \Delta_k$, entonces $\lambda \leq \frac{\Delta_k}{\|\nabla f(x_k)\|}$
- El máximo descenso se obtiene para $\lambda = \frac{\Delta_k}{\|\nabla f(x_k)\|}$, por lo que $p_k^S = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 1)

- Encontrar el punto p_k^S que resuelve la versión lineal:

$$p_k^S = \arg\min_p f(x_k) + \nabla f(x_k)^T p, \text{ s.t. } \|p\| \leq \Delta_k$$

- La función decrece a lo largo de $-\nabla f(x_k)^T$, luego $p_k^S = -\lambda \nabla f(x_k)$ con $\lambda > 0$
- Como $\|p_k^S\| \leq \Delta_k$, entonces $\lambda \leq \frac{\Delta_k}{\|\nabla f(x_k)\|}$
- El máximo descenso se obtiene para $\lambda = \frac{\Delta_k}{\|\nabla f(x_k)\|}$, por lo que $p_k^S = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$
  - $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$
  - $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - Si $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$ entonces $m_k(\tau_k p_k^S)$ decrece a lo largo de p_k^S , i.e., del $-\nabla f(x_k)$, y se toma a τ como el mayor valor posible, es decir $\tau = 1$.

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - Si $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$ entonces $m_k(\tau_k p_k^S)$ decrece a lo largo de p_k^S , i.e., del $-\nabla f(x_k)$, y se toma a τ como el mayor valor posible, es decir $\tau = 1$.

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - Si $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$ entonces $m_k(\tau_k p_k^S)$ es una cuadrática covexa en τ. Si el mínimo se alcanza en el interior de la región de confianza, entonces $\tau = \|\nabla f(x_k)\|^3 / (\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k))$, en caso contrario la solución está en la frontera, $\tau = 1$ similar al caso anterior.

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg \min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- Para hallar una fórmula cerrada para τ_k se consideran 2 casos:
  - Si $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$ entonces $m_k(\tau_k p_k^S)$ es una cuadrática covexa en τ. Si el mínimo se alcanza en el interior de la región de confianza, entonces $\tau = \|\nabla f(x_k)\|^3 / (\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k))$, en caso contrario la solución está en la frontera, $\tau = 1$ similar al caso anterior.

Cálculo del Punto de Cauchy

Algoritmo: Paso a Paso (Paso 2)

- Encontrar el parámetro $\tau_k > 0$ que minimiza $m_k(\tau_k p_k^S)$ en la región de confianza, i.e.,

$$\tau_k = \arg\min_{\tau \geq 0} m_k(\tau p_k^S), \text{ s.t. } \|\tau p_k^S\| \leq \Delta_k$$

- **Resumiendo:** $p_k^C = -\tau_k \frac{\Delta_k}{\|\nabla f_k\|} \nabla f_k$.

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

Representación gráfica: Paso de Cauchy

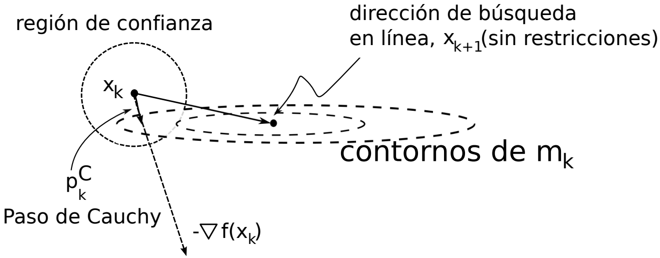

Figura: Punto de Cauchy

Punto de Cauchy: Otra forma de calcularlo

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.

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

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

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

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