Orden del Tema

El Punto de Cauchy es el minimizador del modelo $m_k(\boldsymbol{p})$ a lo largo de de la dirección del máximo descenso, i.e., $\boldsymbol{p}_k = -\lambda_k \boldsymbol{g}_k$ sujeto a la región de confianza.

$$h(\lambda) \quad := \quad m_k(-\lambda \boldsymbol{g}_k) = f_k - \boldsymbol{g}_k^T \boldsymbol{g}_k \lambda + \frac{1}{2}\lambda^2 \boldsymbol{g}_k^T \boldsymbol{B}_k \boldsymbol{g}_k; \ \lambda \geq 0$$

Como $\|\boldsymbol{p}\| \leq \Delta_k$ entonces

$$\| - \lambda \boldsymbol{g}_k\| \leq \Delta_k \ \Rightarrow \lambda \leq \frac{\Delta_k}{\|\boldsymbol{g}_k\|} =: \bar{\lambda}$$

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

Cálculo del Punto de Cauchy

La solución del problema anterior es

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

Resumiendo: $\boldsymbol{p}_k^C = -\lambda_k \boldsymbol{g}_k$. Luego $\boldsymbol{p}_k^C = -\bar{\lambda}\tau_k \boldsymbol{g}_k = -\tau_k \frac{\Delta_k}{\|\boldsymbol{g}_k\|} \boldsymbol{g}_k$

Orden del Tema

Mejorando el Punto de Cauchy

- El Paso de Cauchy $\boldsymbol{p}_k^C$ 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 $\boldsymbol{p}_k^C$ 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 $\boldsymbol{p}_k^C$ 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 $\boldsymbol{p}_k^C$, 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 $\boldsymbol{p}_k^C$ no depende fuertemente de la matriz $\boldsymbol{B}_k$.
- Una mejor convergencia podría encontrarse si se usa la matriz $\boldsymbol{B}_k$.
- Varios algoritmos de Región de Confianza primero calculan $\boldsymbol{p}_k^C$ y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

- El Paso de Cauchy $\boldsymbol{p}_k^C$ no depende fuertemente de la matriz $\boldsymbol{B}_k$.
- Una mejor convergencia podría encontrarse si se usa la matriz $\boldsymbol{B}_k$.
- Varios algoritmos de Región de Confianza primero calculan $\boldsymbol{p}_k^C$ y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

- El Paso de Cauchy $\boldsymbol{p}_k^C$ no depende fuertemente de la matriz $\boldsymbol{B}_k$.
- Una mejor convergencia podría encontrarse si se usa la matriz $\boldsymbol{B}_k$.
- Varios algoritmos de Región de Confianza primero calculan $\boldsymbol{p}_k^C$ y luego tratan de mejorar la solución.

Mejorando el Punto de Cauchy

Observaciones

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

Mejorando el Punto de Cauchy

Observaciones

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

Mejorando el Punto de Cauchy

Observaciones

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

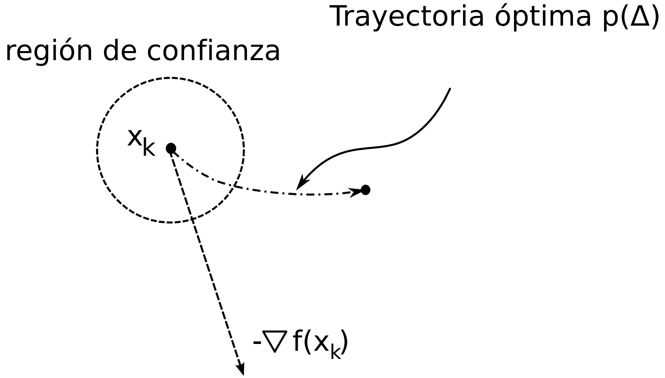
Método Dogleg

Idea del Algoritmo

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

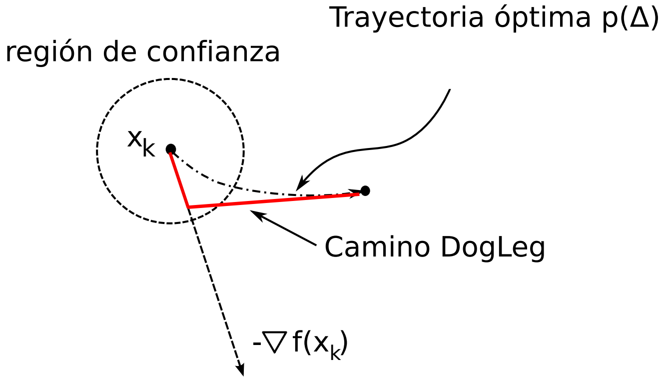
Trayectoria óptima $p(\Delta)$

Cual es el paso óptimo al variar la región de confianza? Si tomamos $\Delta \in [0, \|p_k^B\|]$ la trayectoria óptima $p(\Delta)$ sería



Trayectoria DogLeg

Aproximación de la trayectoria óptima usando el camino Dogleg



Método Dogleg

Idea del Algoritmo

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

Método Dogleg

Trayectoria Dogleg

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

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

Método Dogleg

Minimizar el model cuadrático a lo largo del gradiente: $\boldsymbol{p}_k^U$

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

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

$$\alpha^* = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T \boldsymbol{B}_k \nabla f_k}$$

- Luego $\boldsymbol{p}_k^U = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T \boldsymbol{B}_k \nabla f_k} \nabla f_k$

Método Dogleg

Minimizar el model cuadrático a lo largo del gradiente: $\boldsymbol{p}_k^U$

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

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

$$
\alpha^* = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T \boldsymbol{B}_k \nabla f_k}
$$

- Luego $\boldsymbol{p}_k^U = -\frac{\nabla^T f_k \nabla f_k}{\nabla f_k^T \boldsymbol{B}_k \nabla f_k} \nabla f_k$

Método Dogleg

ww

Luego

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

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

- Como $\boldsymbol{p}_k^U$ y $\boldsymbol{p}_k^B$ son conocidos, entonces lo que tenemos que hallar es τ^*.

Método Dogleg

Lema

Si B es positiva definida entonces:

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

Método Dogleg

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

Método Dogleg

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

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

pues $\boldsymbol{g}_k^T\boldsymbol{B}_k^{-1}\boldsymbol{g}_k\boldsymbol{g}_k^T\boldsymbol{B}_k\boldsymbol{g}_k \geq (\boldsymbol{g}_k^T\boldsymbol{g}_k)^2$, 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 $\tilde{\boldsymbol{p}}(\tau) = \tau \boldsymbol{p}_k^U$

$$
\begin{aligned}
m(\tilde{\boldsymbol{p}}(\tau)) &= f_k + \boldsymbol{g}_k^T \tilde{\boldsymbol{p}}(\tau) + \frac{1}{2}\tilde{\boldsymbol{p}}(\tau)^T \boldsymbol{B}_k \tilde{\boldsymbol{p}}(\tau) \\
m'(\tilde{\boldsymbol{p}}(\tau)) &= (\tilde{\boldsymbol{p}}'(\tau))^T \boldsymbol{g}_k + (\tilde{\boldsymbol{p}}'(\tau))^T \boldsymbol{B}_k \tilde{\boldsymbol{p}}(t) \\
&= (\boldsymbol{p}_k^U)^T \boldsymbol{g}_k + \tau (\boldsymbol{p}_k^U)^T \boldsymbol{B}_k \boldsymbol{p}_k^U \\
&\leq (\boldsymbol{p}_k^U)^T \boldsymbol{g}_k + (\boldsymbol{p}_k^U)^T \boldsymbol{B}_k \boldsymbol{p}_k^U, \text{ tomando } \tau = 1 \\
&\leq \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{g}_k^T \boldsymbol{B}_k \boldsymbol{g}_k}\left[-\boldsymbol{g}_k^T \boldsymbol{g}_k + \frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{g}_k^T \boldsymbol{B}_k \boldsymbol{g}_k} \boldsymbol{g}_k^T \boldsymbol{B}_k \boldsymbol{g}_k\right] = 0
\end{aligned}
$$

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

Método Dogleg

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

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

pues $\boldsymbol{g}_k + \boldsymbol{B}_k \boldsymbol{p}_k^B = 0$ luego $m'(\tilde{\boldsymbol{p}}(t)) \leq 0$ y por tanto $m(\tilde{\boldsymbol{p}}(t))$ es decreciente

Método Dogleg

Observaciones

Basados en el Lema anterior se tiene lo siguiente

- La trayectoria $\tilde{\boldsymbol{p}}(\tau)$ intercepta a la *región de confianza*, $\|\boldsymbol{p}\| = \Delta_k$, en un sólo punto si $\| \boldsymbol{p}_k^B \| \geq \Delta_k$.
- Si $\| \boldsymbol{p}_k^B \| \leq \Delta_k$ entonces el tamaño de paso óptimo es $\boldsymbol{p}_k = \boldsymbol{p}_k^B$, puesto que $m(\tilde{\boldsymbol{p}}(\tau))$ decrece a lo largo del camino Dogleg.

Método Dogleg

Observaciones

- En otro caso, ie, si $\| \boldsymbol{p}_k^B \| > \Delta_k$, hay que hallar el intercepto entre la trayectoria Dogleg y la region de confianza,
  - Si $\|\boldsymbol{p}_k^U\| \geq \Delta_k$ entonces $\boldsymbol{p}_k = \boldsymbol{p}_k^C$
  - De lo contrario (ie, $\|\boldsymbol{p}_k^U\| < \Delta_k$), se tiene que resolver la siguiente ecuación para τ

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

Método Dogleg

A partir de

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

Tenemos que resolver la ecuación cuadrática

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

donde $\tau = \lambda + 1$ y

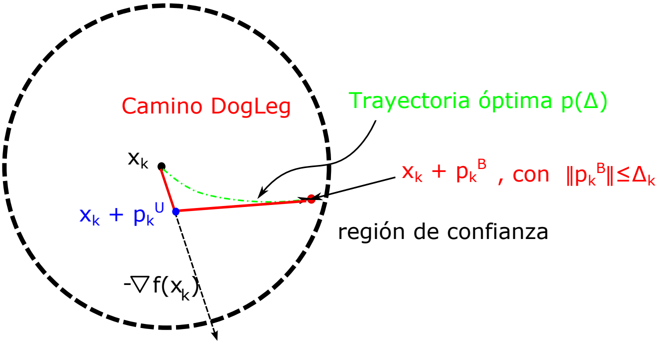
$$a = \|\boldsymbol{p}_k^B - \boldsymbol{p}_k^U\|^2, \; b = 2(\boldsymbol{p}_k^B)^T(\boldsymbol{p}_k^B - \boldsymbol{p}_k^U), \; c = \|\boldsymbol{p}_k^U\|^2 - \Delta^2$$

Finalmente

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

Caso 1: DogLeg

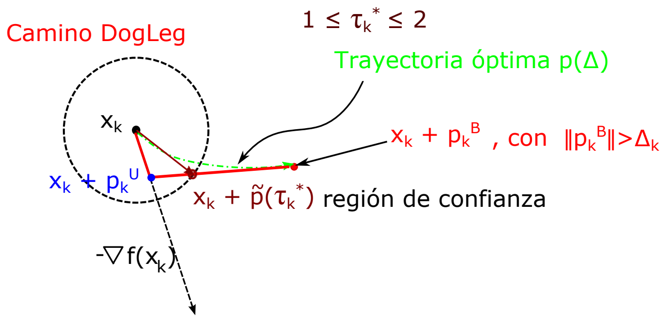
Si p_k^B está en el interior de la región de confianza



$\|p_k^B\| \leq \Delta_k$,i.e. $\|p_k^U\| \leq \Delta_k$ y $p_k^C = p_k^U$

Caso 2: DogLeg

Si p_k^B está fuera de la región de confianza y $1 \leq \tau^* \leq 2$

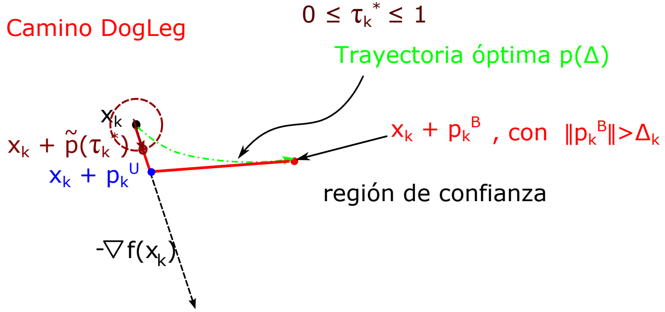


$\|p_k^U\| \leq \Delta_k \Rightarrow p_k^C = p_k^U$

$\|p_k^B\| > \Delta_k$

Caso 3: DogLeg

Si $\boldsymbol{p}_k^B$ está fuera de la región de confianza y $0 \leq \tau^* \leq 1$, ie, $\|\boldsymbol{p}_k^U\| \geq \Delta_k$



$\|p_k^U\| \geq \Delta_k \implies p_k^C = \tau_k^* p_k^U$ y $\|p_k^C\| = \Delta_k$

$\|p_k^B\| > \Delta_k$

Trust region-Dogleg

Given $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \frac{1}{4})$
for $k = 0, 1, 2, \cdots$ **do**
    Compute $\boldsymbol{p}_k$ using Dogleg
    Compute $\rho_k = \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{p}_k)}{m_k(0) - m_k(\boldsymbol{p}_k)}$
    if $\rho_k > \eta$ **then**
        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$
    else
        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$
    end if
    Compute Δ_{k+1} using $(\rho_k, \boldsymbol{p}_k, \Delta_k, \rho_k, \hat{\Delta})$.
end for