Punto de Cauchy. Método *Dogleg*.

Alan Reyes-Figueroa

Métodos Numéricos II (Aula 23) 29.septiembre.2021

Teorema (Caracterización de soluciones para Región de Confianza)

El vector $\mathbf{d}^* \in \mathbb{R}^n$ *es una solución global del problema*

$$\mathbf{d}_k = \text{argmin}_{\mathbf{d}} f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B_k \mathbf{d}, \quad \textit{sujeto a } ||\mathbf{d}|| \leq \Delta_k, \tag{1}$$

$\iff \mathbf{d}^*$ *es factible y existe* $\lambda \geq 0$ *tal que*

$$\begin{aligned} (B + \lambda I)\,\mathbf{d}^* &= -\mathbf{g}, \; (\nabla_{\mathbf{d}}\mathcal{L} = \mathbf{0}) \\ \lambda(\Delta^2 - ||\mathbf{d}^*||^2) &= 0, \quad (\textit{condición de complementariedad}) \\ B + \lambda I &\succeq 0. \end{aligned} \tag{2}$$

Prueba: ($\Leftarrow$) Suponga que existe $\lambda \geq 0$ que satisface las condiciones (2). El lema anterior implica que $\mathbf{d}^*$ es un mínimo global de la función

$$\widetilde{m}(\mathbf{d}) = \mathbf{g}^T\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T(B + \lambda I)\mathbf{d} = m(\mathbf{d}) + \tfrac{1}{2}\lambda\mathbf{d}^T\mathbf{d}.$$

Como $\widetilde{m}(\mathbf{d}) \geq \widetilde{m}(\mathbf{d}^*)$, tenemos que $m(\mathbf{d}) \geq m(\mathbf{d}^*) + \frac{1}{2}\lambda\big((\mathbf{d}^*)^T\mathbf{d}^* - \mathbf{d}^T\mathbf{d}\big)$. Además, de $\lambda(\Delta^2 - ||\mathbf{d}^*||^2) = \lambda(\Delta^2 - ||(\mathbf{d}^*)^T\mathbf{d}^*||^2) = 0$, obtenemos

$$m(\mathbf{d}) \geq m(\mathbf{d}^*) + \tfrac{1}{2}\lambda(\Delta^2 - \mathbf{d}^T\mathbf{d}). \tag{3}$$

Y como $\lambda \geq 0$, resulta $m(\mathbf{d}) \geq m(\mathbf{d}^*)$ para todo $\mathbf{d} \in \mathbb{R}^n$, con $||\mathbf{d}|| \leq \Delta$. Portanto, $\mathbf{d}^*$ es un mínimo global de (1).

($\Rightarrow$) Suponga ahora que $\mathbf{d}^* \in \mathbb{R}^n$ es una solución global de (1). En el caso que $||\mathbf{d}^*|| \leq \Delta$, tendríamos $\nabla m(\mathbf{d}^*) = \mathbf{g} + B\mathbf{d}^* = \mathbf{0}$, $D^2 m(\mathbf{d}^*) = B \succeq 0$, y las condiciones (2) se satisfacen para $\lambda = 0$.

Si $||\mathbf{d}^*|| = \Delta$, entonces $\mathbf{d}^*$ también es solución del problema

$$\min_{\mathbf{d}} m(\mathbf{d}), \quad \text{sujeto a } ||\mathbf{d}|| = \Delta.$$

Aplicando las condiciones de optimalidad, existe $\lambda \in \mathbb{R}$ tal que el Lagrangiano

$$\mathcal{L}(\mathbf{p}, \lambda) = \mathbf{g}^T\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B\mathbf{d} + \tfrac{1}{2}\lambda(\mathbf{d}^T\mathbf{d} - \Delta^2), \tag{4}$$

tiene un punto estacionario en $\mathbf{d}^*$.

Métodos de Región de Confianza

Luego, $\nabla_{\mathbf{d}}\mathcal{L}(\mathbf{d}^*, \lambda) = \mathbf{g} + B\mathbf{d}^* + \lambda\mathbf{d}^* = \mathbf{0}$. De ahí que $(B + \lambda I)\mathbf{d}^* = -\mathbf{g}$.

Como $m(\mathbf{d}) \geq m(\mathbf{d}^*)$ para todo $\mathbf{d} \in \mathbb{R}^n$, con $(\mathbf{d}^*)^{\mathbf{d}^*} = \Delta^2$, entonces para tales vectores vale

$$\widetilde{m}(\mathbf{d}) \geq m(\mathbf{d}^*) + \tfrac{1}{2}\lambda\big((\mathbf{d}^*)^T\mathbf{d}^* - \mathbf{d}^T\mathbf{d}\big),$$

y $\mathbf{g}^T\mathbf{d} + \frac{1}{2}\mathbf{d}^TB\mathbf{d} \geq \mathbf{g}^T\mathbf{d}^* + \frac{1}{2}(\mathbf{d}^*)^TB\mathbf{d}^* + \frac{1}{2}\lambda\big((\mathbf{d}^*)^T\mathbf{d}^* - \mathbf{d}^T\mathbf{d}\big) - (\mathbf{d}^*)^T(B+\lambda I)\mathbf{d} + \frac{1}{2}\mathbf{d}^TB\mathbf{d}$

$$\begin{aligned}
\Longrightarrow\quad & \tfrac{1}{2}\mathbf{d}^TB\mathbf{d} - \tfrac{1}{2}(\mathbf{d}^*)^TB\mathbf{d}^* - (\mathbf{d}^*)^TB\mathbf{d} - \lambda(\mathbf{d}^*)^T\mathbf{d} + (\mathbf{d}^*)^TB\mathbf{d}^* + \lambda(\mathbf{d}^*)^T\mathbf{d}^* - \tfrac{1}{2}\lambda(\mathbf{d}^*)^T\mathbf{d}^* + \\
& +\tfrac{1}{2}\lambda\mathbf{d}^T\mathbf{d} \geq 0 \\
\Longrightarrow\quad & \big(\tfrac{1}{2}\mathbf{d}^TB\mathbf{d} - (\mathbf{d}^*)^TB\mathbf{d} + \tfrac{1}{2}(\mathbf{d}^*)^TB\mathbf{d}^*\big) + \lambda\big(\tfrac{1}{2}\mathbf{d}^T\mathbf{d} - (\mathbf{d}^*)^T\mathbf{d} + \tfrac{1}{2}(\mathbf{d}^*)^T\mathbf{d}^*\big) \geq 0 \\
\Longrightarrow\quad & \tfrac{1}{2}\big(\mathbf{d}^T - \mathbf{d}^*\big)^T\big(B + \lambda I\big)\big(\mathbf{d} - \mathbf{d}^*\big) \geq 0.
\end{aligned}$$

Como el conjunto de direcciones $\{\pm\frac{\mathbf{d}-\mathbf{d}^*}{||\mathbf{d}-\mathbf{d}^*||} : ||\mathbf{d}|| = \Delta\}$ es denso en S^n, esto muestra que $B + \lambda I \succeq 0$.

Por último, falta mostrar que $\lambda \geq 0$.

Métodos de Región de Confianza

Suponga que $\lambda < 0$, son los únicos valores que satisfacen la primera y tercera ecuación de (2). Como $\mathbf{d}^*$ satisface $(B+\lambda I)\mathbf{d}^* = -\mathbf{g}$ y $B + \lambda I \succeq 0$, del lema anterior sabemos que $\mathbf{d}^*$ es un mínimo global de

$$\widetilde{m}(\mathbf{d}) =^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T(B+\lambda I)\mathbf{d}.$$

Luego, de (3), sabemos que $m(\mathbf{d}) \geq m(\mathbf{d}^*) + \frac{1}{2}\lambda(\Delta^2 - \mathbf{d}^T\mathbf{d})$, para todo $\mathbf{d} \in \mathbb{R}^m$, siempre que $||\mathbf{d}|| \geq ||\mathbf{d}^*|| = \Delta$.

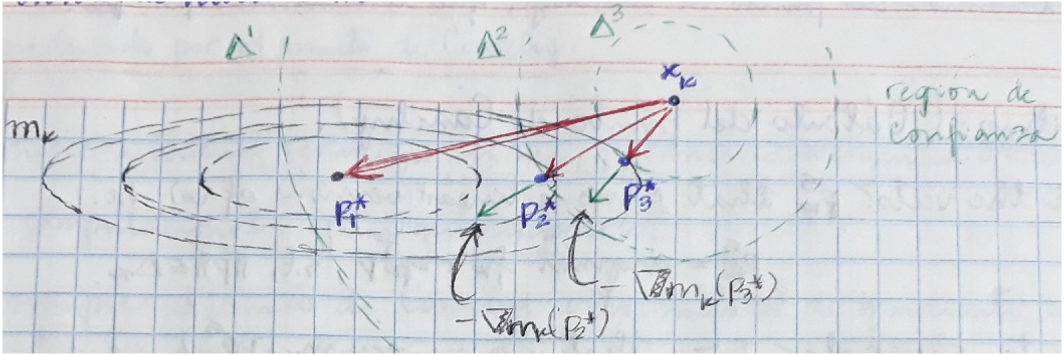
Como ya vimos que $\mathbf{d}^*$ minimiza m en la región $||\mathbf{d}|| \geq \Delta$, se sigue que $\mathbf{d}^*$ es un mínimo global (no restricto) para m, y las condiciones $(B+\lambda I)\mathbf{d}^* = -\mathbf{g}$ y $B + \lambda I \succeq 0$ se satisfacen también para $\lambda = 0$, un absurdo.
Portanto, $\lambda \geq 0$. $\square$

Métodos de Región de Confianza

Comentarios:

- Si $\lambda = 0$ entonces $B\mathbf{d}^* = -\mathbf{g}$. En particular, si $B \succ 0$ entonces $\mathbf{d}^* = -B^{-1}\mathbf{g}$ (*i.e.*, es el paso de Newton o Newton aproximado).
- Si $\lambda > 0$ entonces $B\mathbf{d}^* + \lambda\mathbf{d}^* = -\mathbf{g}$, por lo que $\lambda\mathbf{d}^* = -(B\mathbf{d}^+\mathbf{g})$, es decir, $\mathbf{d}^* = -\frac{1}{\lambda}\nabla m_k(\mathbf{d}^*)$ y por tanto $\mathbf{d}^*$ y $\nabla m_k(\mathbf{d}^*)$ son paralelos, *i.e.*, $\mathbf{d}^*$ es la dirección de máximo descenso del modelo.
- Por otro lado, por la condición, $\lambda(||\mathbf{d}^*|| - \Delta) = 0$ se tiene que cumplir $||\mathbf{d}^*|| = \Delta$, es decir, la solución esta en la frontera (se activa la restricción).

Cuando $\lambda > 0$, la solución $\mathbf{d}^*$ es paralela con el negativo del gradiente $-\nabla m_k(\mathbf{x}_k)$, de modo que es ortogonal a las curvas de nivel de m_k:



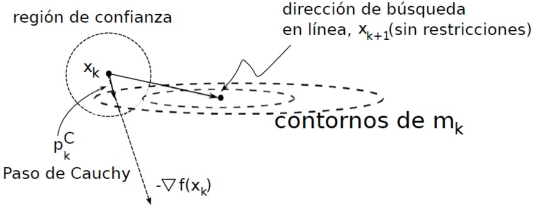
Métodos de Región de Confianza

Vamos a describir una solución que aproxima el subproblema anterior, el cual obtiene una reducción del modelo m_k.
Discutiremos dos estrategias para hallar soluciones aproximadas de (1)

$$\mathbf{d}_k = \text{argmin}_{\mathbf{d}}\, f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B_k \mathbf{d}, \quad \text{sujeto a } ||\mathbf{d}|| \leq \Delta_k,$$

- Punto de Cauchy,
- Método *Dogleg*.

El **punto de Cauchy** de sefine como el minimizador de m_k a lo largo de la dirección de máximo descenso $-\mathbf{g}_k = -\nabla m_k(0) = -\nabla f(\mathbf{x}_k)$, sujeto a la restricción de la región de confianza.



UVG UNIVERSIDAD DEL VALLE DE GUATEMALA