Métodos de Región de Confianza

Alan Reyes-Figueroa

Métodos Numéricos II (Aula 22) 27.septiembre.2021

Métodos de Región de Confianza

Métodos de Región de Confianza: (*Trust region methods*)

Los métodos de región de confianza (MRC), son otra familia de estrategias para optimización. Similar a los métodos de búsqueda en línea, sirven para generar los pasos $\mathbf{x}_k$. Se basan en un modelo que aproxima la función objetivo f, comúnmente una función cuadrática.

Los métodos de región de confianza, 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**.

Idea: Si se encuentra un modelo adecuado de la función objetivo dentro de la región de confianza, entonces la región se expande; por el contrario, si la aproximación es mala, la región se contrae, hasta hallar un modelo apropiado de f.

La función modelo es "confiable" sólo dentro de la región donde proporciona una aproximación razonable.

Métodos de Región de Confianza

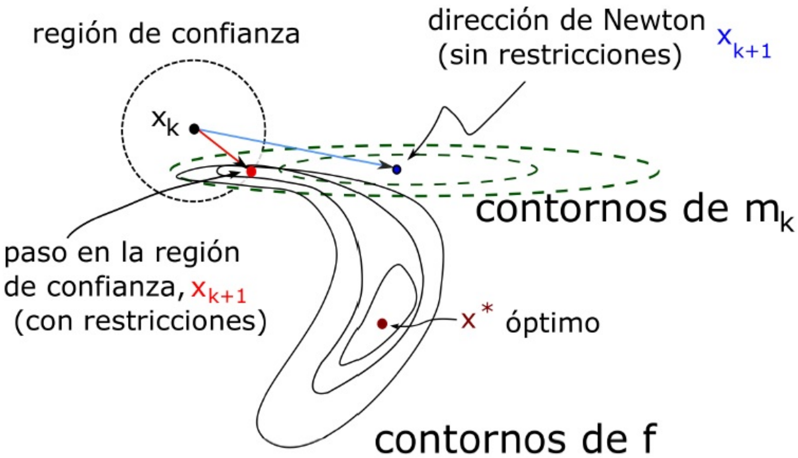
Los métodos de la región de confianza son, en cierto sentido, duales a los métodos de búsqueda de línea:

- los métodos de la región de confianza primero eligen un tamaño de paso α_k y luego una dirección de paso $\mathbf{d}_k$,
- mientras que los métodos de búsqueda de línea primero eligen una dirección de paso y luego el tamaño de paso.

La idea general detrás de los métodos de región de confianza se conoce por muchos nombres

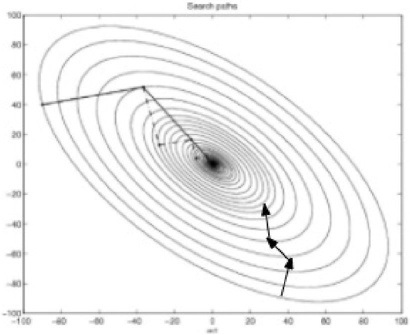
- El primer uso del término aparece con SORENSEN, "Newton's Method with a Model Trust Region Modification". SIAM J. Numer. Anal. 19 (2): 409—426 (1982).
- FLETCHER (1980) llama a estos algoritmos métodos de pasos restringidos en "Restricted Step Methods". Practical Methods of Optimization (2a ed.). Wiley.
- GOLDFELD, QUANDT y TROTTER (1966) se refieren a él como escalada cuadrática (*quadratic hill-climbing*).

Métodos de Región de Confianza

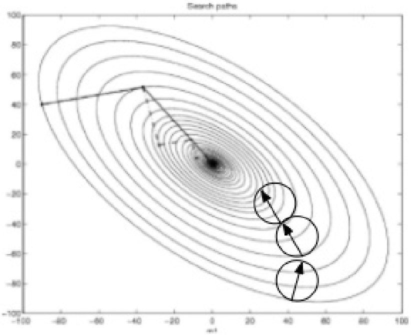


Comparación de pasos calculados usando región de confianza y método de Newton.

Métodos de Región de Confianza



LINE SEARCH METHOD



TRUST REGION METHOD

Métodos de Región de Confianza

- El paso se calcula como un minimizador aproximado de un modelo, $m_k(\mathbf{d})$, restricto a una región de confianza $\mathcal{R}_k$.
- 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.

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k.$$

- 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 o radio de la región y se recalcula el paso usando la nueva región de confianza.

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.

Ingredientes para los algoritmos de región de confianza:

- Un modelo $m_k(\mathbf{d})$,
- Radio o tamaño de la región de confianza Δ_k,
- Una medida para evaluar el ajuste del modelo en la región de confianza.

Considere $f : \mathbb{R}^n \to \mathbb{R}$ una función de clase C^2. Sea $\{\mathbf{x}_k\} \subset \mathbb{R}^n$ una secuencia de aproximaciones al mínimo de f

$$\mathbf{x}^* = \operatorname{argmin} f(\mathbf{x}),$$

y sea $\mathbf{d} \in \mathbb{R}^n$ una dirección (de descenso) en $\mathbf{x}_k$.

Asumimos un modelo cuadrático $m_k : \mathbb{R}^n \to \mathbb{R}$ que aproxima a f con base en la serie de Taylor

$$f(\mathbf{x}_k + \mathbf{d}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T D^2 f(\mathbf{x}_k + t\mathbf{d})\mathbf{d},$$

donde $t \in (0, 1)$.

Usando una aproximación B_k del hessiano $D^2 f(\mathbf{x}_k + t\mathbf{d}_k)$, el modelo m_k se define como

$$m_k(\mathbf{d}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B_k \mathbf{d}, \tag{1}$$

con B_k simétrica.

La diferencia entre $f(\mathbf{x}_k + \mathbf{d})$ y $m_k(\mathbf{d})$ es de orden $O(||\mathbf{d}||^2)$. Cuando $B_k = D^2 f(\mathbf{x}_k)$ es el hessiano eacto, la diferencia es de orden $O(||\mathbf{d}||^3)$.

Métodos de Región de Confianza

- El caso $B_k = D^2 f(\mathbf{x}_k)$ da origen al método de región de confianza de **Newton**.
- En el caso de métodos de región de confianza en general, sólo se asume que B_k es simétrica y uniformemente limitada ($\exists M > 0$ con $||B_k|| < M$, $\forall k$).

Para obtener el paso $\mathbf{x}_k$, los métodos de región de confianza resuelven el problema de optimización restricta

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

donde $\Delta_k > 0$ es el **radio** de la región de confianza. Observe que la restricción puede escribirse como $\mathbf{d}^T\mathbf{d} \leq \Delta_k^2$, asi, el método requiere resolver un problema donde la función objetivo y la restricción son cuadráticos.

Cuando B_k es positiva definida y $||B_k \nabla f(\mathbf{x}_k)|| \leq \Delta_k$, la solución de (2) es simplemente el mínimo sin restricción:

$$\mathbf{d}_k = \text{argmin}_{\mathbf{d}}\, m_k(\mathbf{d}) = -B_k \nabla f(\mathbf{x}_k). \tag{3}$$

Métodos de Región de Confianza

La solución en (3) se llama la solución de **paso completo** (*full-step*).

En cualquier otro caso, la solución del problema

$$\mathbf{d}_k = \text{argmin}_{\mathbf{d}} f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B_k \mathbf{d}, \quad \text{sujeto a } ||\mathbf{d}|| \leq \Delta_k, \tag{4}$$

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.

UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Calidad del Modelo

Un elemento importante en los métodos de región de confianza es la forma de calcular el radio de la región Δ_k en cada iteración.
Hacemos esta elección con base en el acuerdo entre la función del modelo m_k y la función objetivo f, en la iteración previa. Para ello, definimos la siguiente medida del ajuste

$$\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{d}_k)}{m_k(\mathbf{0}) - m_k(\mathbf{d}_k)}. \tag{5}$$

El numerador en (5) se llama la **reducción real**, el denominador es la **reducción predicha**.

Cuando $\mathbf{d}_k$ se obtiene minimizando el modelo m_k sobre una región que incluye al cero $\mathbf{d} = \mathbf{0}$, se tiene que $m_k(\mathbf{0}) - m_k(\mathbf{d}_k) \geq 0$.

- Así, si $\rho_k < 0$, se tiene $f(\mathbf{x}_k + \mathbf{d}_k) > f(\mathbf{x}_k)$, y el paso debe rechazarse.
- Por otro lado, si $\rho_k \approx 1$, hay un buen acuerdo ente el modelo m_k y la función f, de modo que es seguro expandir la región de confianza en la próxima iteración.
- Si $0 < \rho_k < 1$, no se modifica el radio en la próxima iteración, i.e., $\Delta_{k+1} = \Delta_k$.
- Si $\rho_k \approx 0$ ó $\rho_k < 0$, la región de confianza se reduce Δ_k .

Métodos de Región de Confianza

Algoritmo: (*Región de Confianza*)

Inputs: $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^2, $m : \mathbb{R}^n \to \mathbb{R}$ modelo aproximadose, $\mathbf{x}_k \in \mathbb{R}^n$, $\widehat{\Delta} > 0$ radio máximo, $\Delta_0 \in (0, \widehat{\Delta})$ radio inicial, $\eta \in [0, \frac{1}{4})$ parámetro de control.

```
For k = 0, 1, 2, ...:
Obtain descent direction d_k by approximately solving (4), and compute ρ_k,
    If (ρ_k < 1/4) :
        Compute Δ_{k+1} = 1/4 Δ_k,
    Else:
        If (ρ_k > 3/4) and (||d_k|| == Δ_k):
            Δ_{k+1} = min{2Δ_k, Δ̂},
        Else:
            Δ_{k+1} = Δ_k.
    If (ρ_k > η): x_{k+1} = x_k + d_k,
    Else: x_{k+1} = x_k,
Return x_{k+1}.
```

Algoritmo 2: (*Región de Confianza*)
Inputs: $f : \mathbb{R}^n \to \mathbb{R}$ de clase C^2, $m : \mathbb{R}^n \to \mathbb{R}$ modelo aproximadose, $\mathbf{x}_k \in \mathbb{R}^n$, $\widehat{\Delta} > 0$ radio máximo, $\Delta_0 \in (0, \widehat{\Delta})$ radio inicial, $\eta_1, \eta_2, \widehat{\eta}_1, \widehat{\eta}_2 \in [0,1)$, $\eta \in [0, \eta_1]$ parámetros de control.

```
For k = 0, 1, 2, ...:
Obtain descent direction d_k by approximately solving (4), and compute ρ_k,
    If (ρ_k < η_1) :
        Compute Δ_{k+1} = η̂_1 Δ_k,
    Else:
        If (ρ_k > η_2) and (||d_k|| == Δ_k):
            Δ_{k+1} = min{η̂_2 Δ_k, Δ̂},
        Else: # Si ρ_k ∈ [η_1, η_2] ó ||d_k|| < Δ_k
            Δ_{k+1} = Δ_k.
    If (ρ_k > η): x_{k+1} = x_k + d_k,
    Else: x_{k+1} = x_k,
Return x_{k+1}.
```

Métodos de Región de Confianza

Parámetros por defecto: $\eta_1 = \frac{1}{4}$, $\eta_2 = \frac{3}{4}$, $\widehat{\eta}_1 = \frac{1}{4}$, $\widehat{\eta}_2 = 2$.

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 ρ_k alcanza el borde de la región de confianza, i.e., $||\rho_k|| = \Delta_k$.
- 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 sólo se reduce, si la función incrementa su valor, o si el modelo y la función no concuerdan bien, *i.e.*, $\rho_k < \frac{1}{4}$.

Métodos de Región de Confianza

¿Qué falta para completar el algoritmo?

¿Cómo calcular del Paso?
Para hallar el paso, se resuelve el siguiente problema de optimimización con restricciones (4):

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

Aquí, Δ_k es el radio de la región de confianza. Como se comentó anteriormente, la solución del problema anterior puede ser muy compleja.

Lema

Sea $m : \mathbb{R}^n \to \mathbb{R}$ la función cuadrática definida por $m(\mathbf{d}) = \mathbf{g}^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T B\mathbf{d}$, con B simétrica. Valen las siguientes:

i) *m posee un mínimo $\iff$ B es positiva semi-definida y $\mathbf{g} \in \operatorname{Im} B$.*
Si $B \succeq 0$, todo $\mathbf{d}$ que satisface $B\mathbf{d} = -\mathbf{g}$ es un mínimo global de m.

ii) *m posee un único mínimo local $\iff$ B es positiva definida.*

UNIVERSIDAD DEL VALLE DE GUATEMALA

Métodos de Región de Confianza

Prueba: (i) ($\Leftarrow$) Como $\mathbf{g} \in \operatorname{Im} B$, existe $\mathbf{d} \in \mathbb{R}^n$ tal que $B\mathbf{d} = -\mathbf{g}$. Luego, $\forall \mathbf{w} \in \mathbb{R}^n$ vale

$$
\begin{aligned}
m(\mathbf{d}+\mathbf{w}) &= \mathbf{g}^T(\mathbf{d}+\mathbf{w}) + \tfrac{1}{2}(\mathbf{d}+\mathbf{w})^T B(\mathbf{d}+\mathbf{w}) \\
&= \left(\mathbf{g}^T\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T B\mathbf{d}\right) + \left(\mathbf{g}^T\mathbf{w} + (B\mathbf{d})^T\mathbf{w} + \tfrac{1}{2}\mathbf{w}^T B\mathbf{w}\right) \\
&= m(\mathbf{d}) + \left(\mathbf{g}^T\mathbf{w} - \mathbf{g}^T\mathbf{w} + \tfrac{1}{2}\mathbf{w}^T B\mathbf{w}\right) = m(\mathbf{d}) + \tfrac{1}{2}\mathbf{w}^T B\mathbf{w} \\
&\geq m(\mathbf{d}),
\end{aligned}
$$

pues $B \succeq 0$. De ahí que $\mathbf{d}$ es un mínimo global de m.

($\Rightarrow$) Sea $\mathbf{d} \in \mathbb{R}^n$ un mínimo de m. Entonces, $\nabla m(\mathbf{d}) = \mathbf{g} + B\mathbf{d} = \mathbf{0} \quad B\mathbf{d} = -\mathbf{g}$. Así que $\mathbf{g} \in \operatorname{Im} B$. Además, $D^2 m(\mathbf{d}) = B \succeq 0$ es positiva semi-definida. La segunda afirmación es consecuencia de las condiciones de optimalidad.

(ii) ($\Leftarrow$) Para todo $\mathbf{w} \in \mathbb{R}^n$, el mismo argumento usado en (i) implica que

$$
m(\mathbf{d}+\mathbf{w}) = m(\mathbf{d}) + \tfrac{1}{2}\mathbf{w}^T B\mathbf{w} > m(\mathbf{d}),
$$

pues $B \succ 0$, siempre que $\mathbf{w} \neq 0$. Así, $\mathbf{d}$ es el único mínimo de m.

Métodos de Región de Confianza

($\Rightarrow$) Sabemos que $\nabla m(\mathbf{d}) = \mathbf{g} + B\mathbf{d} = \mathbf{0} \;\Rightarrow\; B\mathbf{p} = -\mathbf{g}$. Si B no fuese positiva definida, existe $\mathbf{w} \neq \mathbf{0}$ tal que $\frac{1}{2}\mathbf{w}^T B\mathbf{w} \leq 0$ (más concretamente, $\frac{1}{2}\mathbf{w}^T B\mathbf{w} = 0$. De ahí que

$$m(\mathbf{d}+\mathbf{w}) = m(\mathbf{d}) + \underbrace{\tfrac{1}{2}\mathbf{w}^T B\mathbf{w}}_{=0} = m(\mathbf{d}),$$

y el mínimo no sería único. Portanto, $B \succ 0$. $\square$

Un primer paso para caracterizar las soluciones del problema (4)

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

está dado por el siguiente resultado de Moré y Sorensen.

Teorema (Caracterización de soluciones para Región de Confianza)

El vector $\mathbf{d}^* \in \mathbb{R}^n$ es una solución global del problema (4) $\Longleftrightarrow$ $\mathbf{d}^*$ es factible y existe $\lambda \geq 0$ tal que

Métodos de Región de Confianza

$$
\begin{aligned}
(B + \lambda I)\,\mathbf{d}^* &= -\mathbf{g}, \; (\nabla_{\mathbf{d}}\mathcal{L} = \mathbf{0}) \\
\lambda(\Delta^2 - ||\mathbf{d}^*||^2) &= 0, \quad \text{(condición de complementariedad)} \\
B + \lambda I &\succeq 0.
\end{aligned}
\tag{6}
$$

UNIVERSIDAD DEL VALLE DE GUATEMALA