CIMAT

Centro de Investigación en Matemáticas, A.C.
Optimización

Examen Parcial 1

José Miguel Saavedra Aguilar

Problema 1

Considere el algoritmo

$$x_{k+1} = x_k - \alpha_k H_k g_k \tag{1}$$

donde $f : \mathbb{R}^3 \to \mathbb{R}, f \in \mathcal{C}^2(\mathbb{R}^3).g_k = \nabla f(x_k)$ y $H_k \in \mathbb{R}^{3\times 3}$ es una matriz dada como sigue

$$H_k = \begin{bmatrix} a^2 & -a & -a \\ -a & -a^2 & 3a \\ -a & 3a & -a^2 \end{bmatrix} \tag{2}$$

con $a \in \mathbb{R}$. Suponga que en la iteración k-ésima del algoritmo se cumple que $g_k = [1, 1, 1]^\top$. Encuentre los valores de $a \in \mathbb{R}$ para los cuales existe $\alpha_k > 0$ que satisface:

$$\alpha_k = \underset{\alpha > 0}{\arg\ \min}\ \ f(x_k - \alpha H_k g_k) \tag{3}$$

Para garantizar que α_k pueda existir, primero debemos asegurar que $d_k = -H_k g_k$ sea una dirección de descenso. Por lo tanto, consideramos:

$$g_k^\top d_k = -g_k^\top H_k g_k$$
$$g_k^\top d_k = -(2a - a^2)$$
$$g_k^\top d_k = a(a - 2)$$

Y para asegurar $g_k^\top d_k < 0$, lo que se garantiza siempre que $a \in (0, 2)$. La existencia de α_k también depende de condiciones sobre f, sin embargo no imponen condiciones sobre a, por lo que tenemos que $a \in (0, 2)$ es lo que requerimos asegurar sobre a.

Problema 2

Sea $Q \in \mathbb{R}^{2\times 2}$ simétrica con eigenvalores $0 < \lambda_1 < \lambda_2$ y sean u_1, u_2 los eigenvectores asociados a λ_1, λ_2, respectivamente.
Aplique el método de descenso máximo con búsqueda en línea exacta para minimizar la función

$$f(x) = \frac{1}{2} x^\top Q x \tag{4}$$

partiendo del punto

$$x_0 = \frac{1}{\lambda_1} u_1 + \frac{1}{\lambda_2} u_2 \tag{5}$$

a)

Encuentre la expresión para x_k.
Nótese que x_{k+1} puede ser expresado por:

$$x_{k+1} = U \left[I - \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} \Lambda \right] v_k$$

Donde $Q = U \Lambda U^\top$ es una matriz diagonal con entradas λ_1, λ_2 y $U = [u_1, u_2]$ y $v_k = U^\top x_k$.
De forma que x_1 está dado por:

$$x_1 = U \left[I - \frac{v_0^\top \Lambda^2 v_0}{v_0^\top \Lambda^3 v_0} \Lambda \right] v_0$$
$$x_1 = U \left[I - \frac{2}{\lambda_1 + \lambda_2} \Lambda \right] v_0$$

Además, notamos que

$$\frac{v_1^\top \Lambda^2 v_1}{v_1^\top \Lambda^3 v_1} = \frac{2}{\lambda_1 + \lambda_2}$$

De forma que demostraremos por inducción

$$x_k = U \left[I - \frac{2}{\lambda_1 + \lambda_2} \Lambda \right]^k v_0 \tag{6}$$

Los casos $k = 0, 1, 2$ ya se tienen, por lo que supongamos que x_k cumple (6). Sea x_{k+1}:

$$x_{k+1} = U \left[I - \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} \Lambda \right] v_k$$
$$x_{k+1} = U \left[I - \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} \Lambda \right] U^\top x_k$$
$$x_{k+1} = U \left[I - \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} \Lambda \right] \left[I - \frac{2}{\lambda_1 + \lambda_2} \Lambda \right]^k v_0$$

Una vez que se cumple

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{(\left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^k v_0)^\top \Lambda^2 \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^k v_0}{(\left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^k v_0)^\top \Lambda^3 (\left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^k v_0)}$$

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{(\Lambda v_0)^\top \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^{2k} \Lambda v_0}{(\Lambda v_0)^\top \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^{2k} \Lambda^2 v_0}$$

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{[1,1] \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^{2k} [1,1]^\top}{[1,1] \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda\right]^{2k} [\lambda_1, \lambda_2]^\top}$$

Una vez que $I - \frac{2}{\lambda_1+\lambda_2}\Lambda$ es diagonal, elevarla a la $2k$ es equivalente a elevar cada elemento de su diagonal a la $2k$:

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{\left(1 - \frac{2\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} + \left(1 - \frac{2\lambda_2}{\lambda_1+\lambda_2}\right)^{2k}}{\left(1 - \frac{2\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} \lambda_1 + \left(1 - \frac{2\lambda_2}{\lambda_1+\lambda_2}\right)^{2k} \lambda_2}$$

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{\left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} + \left(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}\right)^{2k}}{\left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} \lambda_1 + \left(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2}\right)^{2k} \lambda_2}$$

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{\left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} + \left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k}}{\left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} \lambda_1 + \left(\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}\right)^{2k} \lambda_2}$$

$$\frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k} = \frac{2}{\lambda_1 + \lambda_2}$$

De forma que

$$x_{k+1} = U \left[I - \frac{2}{\lambda_1 + \lambda_2}\Lambda\right] \left[I - \frac{2}{\lambda_1 + \lambda_2}\Lambda\right]^k v_0$$

$$x_{k+1} = U \left[I - \frac{2}{\lambda_1 + \lambda_2}\Lambda\right]^{k+1} v_0$$

Por lo que (6) se cumple para toda $k \in \mathbb{N}$.

b)

Muestre que x_k tiende al óptimo de f cuando k tiende a infinito.
Recordemos

$$x_{k+1} = U \left[I - \frac{2}{\lambda_1 + \lambda_2}\Lambda\right] U^\top x_k$$

Nótese

$$
\begin{aligned}
\|x_{k+1}\|_2 &= \left\| U \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda \right] U^\top x_k \right\|_2 \\
\|x_{k+1}\| &\leq \left\| U \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda \right] U^\top \right\|_2 \|x_k\|_2
\end{aligned}
$$

Una vez que U es ortogonal, $\left\| U \left[I - \frac{2}{\lambda_1+\lambda_2}\Lambda \right] U^\top \right\|_2 = \left\| I - \frac{2}{\lambda_1+\lambda_2}\Lambda \right\|_2$. Una vez que los eigenvalores de $I - \frac{2}{\lambda_1+\lambda_2}\Lambda$ son $\pm\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2}$, como $0 < \lambda_1 < \lambda_2$, $\frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2} < 1$, de forma que $\left\| I - \frac{2}{\lambda_1+\lambda_2}\Lambda \right\|_2 < 1$. Así:

$$
\|x_{k+1}\| < \|x_k\|_2
$$

De forma que $\|x_k\| \to 0$ cuando $k \to \infty$. Luego, $x_k \to 0$ cuando $k \to \infty$.

c)

Pruebe que

$$
\frac{f(x_{k+1})}{f(x_k)} < \frac{\lambda_2-\lambda_1}{\lambda_1+\lambda_2} \tag{7}
$$

Nótese para $x \neq 0$:

$$
\begin{aligned}
0 \leq \frac{1}{\lambda_2} \leq \frac{x^\top \Lambda^{-1} x}{x^\top x} \leq \frac{1}{\lambda_1} \\
-\lambda_2 \leq \frac{-x^\top x}{x^\top \Lambda^{-1} x} \leq -\lambda_1 \leq 0
\end{aligned}
$$

Además, $\frac{-x^\top x}{x^\top \Lambda^{-1} x} \leq -\lambda_1$ únicamente cuando $x_2 = 0$. Ahora, considere:

$$
\begin{aligned}
\frac{f(x_{k+1})}{f(x_k)} &= \frac{\frac{1}{2} v_k^\top [I - \alpha_k \Lambda] \Lambda [I - \alpha_k \Lambda] v_k}{\frac{1}{2} v_k^\top \Lambda v_k} \\
\frac{f(x_{k+1})}{f(x_k)} &= \frac{\frac{1}{2} v_k^\top v_k + \frac{1}{2}\alpha_k^2 v_k^\top \Lambda^3 v_k - \alpha_k v_k^\top \Lambda^2 v_k}{\frac{1}{2} v_k^\top \Lambda v_k}
\end{aligned}
$$

Tenemos que $\alpha_k = \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda^3 v_k}$, de forma que

$$
\begin{aligned}
\frac{f(x_{k+1})}{f(x_k)} &= \frac{\frac{1}{2} v_k^\top v_k - \frac{1}{2}\alpha_k v_k^\top \Lambda^2 v_k}{\frac{1}{2} v_k^\top \Lambda v_k} \\
\frac{f(x_{k+1})}{f(x_k)} &= 1 - \alpha_k \frac{v_k^\top \Lambda^2 v_k}{v_k^\top \Lambda v_k} \\
\frac{f(x_{k+1})}{f(x_k)} &= 1 - \alpha_k \frac{(\Lambda v_k)^\top (\Lambda v_k)}{(\Lambda v_k)^\top \Lambda^{-1} (\Lambda v_k)}
\end{aligned}
$$

Una vez que $(v_0)_2 \neq 0$, y como $[I - \alpha_k \Lambda]^k$ es diagonal de rango completo, $(v_k)_2 \neq 0$. Luego

$$\begin{aligned}
\frac{f(x_{k+1})}{f(x_k)} &< 1 - \alpha_k \lambda_1 \\
\frac{f(x_{k+1})}{f(x_k)} &< 1 - \frac{2\lambda_1}{\lambda_1 + \lambda_2} \\
\frac{f(x_{k+1})}{f(x_k)} &< \frac{\lambda_2 - \lambda_1}{\lambda_1 + \lambda_2}
\end{aligned}$$

Problema 3

El teorema 4.1 de [1] da la condición de optimalidad del subproblema que aparece en los métodos de región de confianza, ie

$$\begin{aligned}
d_k = \ & \arg \min_p \quad m(p) = f_k + g_k^\top p + \frac{1}{2} B_k \\
& \text{s.t.} \qquad \|p\| \leq \Delta_k
\end{aligned} \tag{8}$$

Entonces d_k es solución de 8 si y solo si existe $\lambda \geq 0$ tal que

$$\begin{aligned}
(B_k + \lambda I) d_k &= -g_k \\
\lambda(\Delta_k - \|d_k\|) &= 0 \\
B_k + \lambda I &\succeq 0
\end{aligned}$$

Si $\lambda = 0$, $d_k = -B_k^{-1} g_k$ es la solución exacta del problema. En otro caso, definimos:

$$p(\lambda) := -(B_k + \lambda I)^{-1} g_k$$

de forma que buscamos la solución de la ecuación

$$\|p(\lambda)\| = \Delta$$

con respecto a λ.
Sin embargo, en la práctica es mejor resolver

$$\phi(\lambda) = 0$$

Para la función $\phi : \mathbb{R} \to \mathbb{R}$ dada por:

$$\phi(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}$$

Podemos calcular la raíz de ϕ utilizando el método de Newton

$$\lambda k + 1 = \lambda_k - \frac{\phi(\lambda_k)}{\phi'(\lambda_k)} \tag{9}$$

(a)

Calcule la derivada de $\|p(\lambda)\|$ respecto a λ.
Por regla de la cadena, tenemos:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\|p(\lambda)\| = -\frac{p(\lambda)^\top p'(\lambda)}{\|p(\lambda)\|}$$

Nótese

$$\begin{aligned}
(B+\lambda I)p(\lambda) &= -g \\
p(\lambda) + (B+\lambda I)p'(\lambda) &= 0 \\
p'(\lambda) &= -(B+\lambda I)^{-1}p(\lambda)
\end{aligned}$$

De forma que:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\|p(\lambda)\| = \frac{p(\lambda)^\top (B+\lambda I)^{-1} p(\lambda)}{\|p(\lambda)\|}$$

(b)

Calcule la derivada de $\phi(\lambda)$.
Sabemos que

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\phi(\lambda) &= \frac{p(\lambda)^\top p'(\lambda)}{\|p(\lambda)\|^3} \\
\frac{\mathrm{d}}{\mathrm{d}\lambda}\phi(\lambda) &= -\frac{p(\lambda)^\top (B+\lambda I)^{-1} p(\lambda)}{\|p(\lambda)\|^3}
\end{aligned}$$

(c)

Muestre que el algoritmo de Newton (9) se puede implementar mediante el Algoritmo 1.
Nótese que

$$\begin{aligned}
\lambda k+1 &= \lambda_k - \frac{\phi(\lambda_k)}{\phi'(\lambda_k)} \\
\lambda k+1 &= \lambda_k - \frac{\frac{1}{\Delta} - \frac{1}{\|p_k\|}}{-\frac{p_k^\top (B+\lambda_k I)^{-1} p_k}{\|p_k\|^3}} \\
\lambda k+1 &= \lambda_k + \frac{\|p_k\|^2 (\|p_k\| - \Delta)}{\Delta p_k^\top (B+\lambda_k I)^{-1} p_k} \\
\lambda k+1 &= \lambda_k + \frac{\|p_k\|^2}{p_k^\top (B+\lambda_k I)^{-1} p_k} \frac{\|p_k\| - \Delta}{\Delta}
\end{aligned}$$

Ahora, suponga que $B + \lambda I$ es SPD, entonces existe R tal que $B + \lambda I = R^\top R$. Además, $(B+\lambda I)^{-1} = R^{-1}R^{-\top}$. Así, resolver $(B+\lambda I)p_k = -g$ es equivalente a resolver $R^\top R p_k = -g$ y $p_k^\top (B + \lambda I)^{-1} p_k = R^{-1}R^{-\top} p_k$, así si q_k es la solución del sistema $R^\top q_k = p_k$, $p_k^\top (B + \lambda I)^{-1} p_k = q_k^\top q_k$. Luego, (9) es equivalente a

$$\lambda k + 1 = \lambda_k + \frac{\|p_k\|^2}{\|q_k\|^2} \frac{\|p_k\| - \Delta}{\Delta}$$

Algorithm 1: Exact trust region.

Input: Δ, g, B
Output: p

for $k = 0, 1, 2, \ldots$ **do**
  Factorizar $B + \lambda_k I = R_k^\top R_k$;
  Resolver $R_k^\top R_k p_k = -g$;
  Resolver $R_k^\top p_k$;
  Actualizar $\lambda_{k+1} = \lambda_k + \frac{\|p_k\|^2}{\|q_k\|^2} \frac{\|p_k\| - \Delta}{\Delta}$;
end

(d)

Implemente el Algoritmo 1 y compare con el algoritmo Dogleg usando la función de Rosenborck en dimensión $n = 2$ como en la tarea 8.

Se implementó una variación del algoritmo 1 sin la factorización de Cholesky de $B + \lambda_k I$ puesto que no podemos garantizar $\lambda_k > 0$ y para ciertos valores de λ_k, $B + \lambda_k I$ no es positiva definida y no tiene factorización de Cholesky, esto es incluso mencionado en [1]. Así mismo, basado en [1] tomamos 3 iteraciones de Newton para aproximar d_k, por lo que en la práctica no hay gran diferencia entre hacer $B + \lambda_k I = R^\top R$ o no . Debemos notar que este algoritmo es inestable numéricamente cuando $\Delta \approx 0$. Por este motivo, tomamos

$$(B + \lambda_k I)p_k = \frac{-1}{\Delta} g$$
$$(B + \lambda_k I)q_k = p_k$$
$$\lambda_{k+1} = \lambda_k + \frac{\|p_k\|^2}{q_k^\top p_k}(\|p_k\| - 1)$$

Con este algoritmo, presentamos la comparativa con el algoritmo de Dogleg:

	Dogleg	Exacta
Iteraciones promedio	14.233	13.366
Tiempo promedio	0.00740	0.01013
Convergencia	30	30

Cuadro 1: Comparativo entre Dogleg y solución exacta

Notamos que no hay gran diferencia entre ambos algoritmos en esta función, sin embargo, a pesar que Dogleg hace en promedio una iteración más que la solución exacta, obtener la solución exacta es más tardado incluso haciendo pocas iteraciones de Newton, por lo que no será la mejor opción en algunos casos.

Referencias

[1] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer, Jul. 2006.