Análisis Aplicado

Dr. Andreas Wachtel

Tarea 4

Algunas tareas son del Dr. Zeferino Parada.

1. Introducción

Sea $f\colon \mathbb{R}^n \to \mathbb{R}$ tal que $f \in C^2(\mathbb{R}^n)$. Supongamos que $\boldsymbol{x}_k \in \mathbb{R}^n$ satisface que

$$\boldsymbol{g}_k \overset{\text{def}}{=\!=} \nabla f(\boldsymbol{x}_k) \neq \boldsymbol{0} \qquad \text{y} \qquad B_k \overset{\text{def}}{=\!=} \nabla^2 f(\boldsymbol{x}_k) \quad \text{es simétrica}\,.$$

En una vecindad de confianza (de $\boldsymbol{x}_k$) con radio $\Delta > 0$ aproximamos a f por un modelo cuadrático. De este modelo buscamos un mínimo, lo cual resulta en el subproblema de región de confianza:

$$\text{(SPRC)} \qquad \left[\begin{array}{rl} \text{Minimizar} & m(\boldsymbol{p}) \overset{\text{def}}{=\!=} f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T \boldsymbol{p} + \dfrac{1}{2}\boldsymbol{p}^T B_k \boldsymbol{p} \\ \text{Sujeto a} & \|\boldsymbol{p}\|_2 \leq \Delta \end{array}\right.$$

De tareas anteriores sabemos que la dirección de Newton $\boldsymbol{p}_N = -B_k^{-1}\boldsymbol{g}_k$ es la solución óptima del (SPRC) cuando B_k es s.p.d. y $\left\|B_k^{-1}\boldsymbol{g}_k\right\|_2 \leq \Delta$. Cuando eso no es cierto, el Teorema (SPRC) dice que para cada radio $\Delta > 0$ existe una solución óptima que depende de un *shift* $s \geq 0$ de la siguiente forma:

$$\text{(1)} \qquad \boldsymbol{p}(s) = -(B_k + sI)^{-1}\boldsymbol{g}_k \qquad \text{y} \qquad s(\Delta - \|\boldsymbol{p}(s)\|_2) = 0\,,$$

con $(B + sI)$ simétrica positiva semidefinida. Si el radio Δ cambia, entonces s cambia, es decir, s depende de Δ pero no denotamos esa dependencia. La perspectiva opuesta es ver como cambia $\Delta(s) \overset{\text{def}}{=\!=} \|\boldsymbol{p}(s)\|_2$ si aumentamos a s. Se puede ver que $\boldsymbol{p}(s)$ para $s \in [0, \infty)$ define una curva con $\boldsymbol{p}(s) \to \boldsymbol{0}$ cuando $s \to \infty$.

Además, hemos definido el punto de Cauchy que es un múltiplo de $-\boldsymbol{g}$.

2. Problemas

1. Sea $\boldsymbol{p}_\star$ la solución del (SPRC) tal que $\boldsymbol{g}_k^T \boldsymbol{p}_\star \neq 0$.
Demuestre que $\nabla f(\boldsymbol{x}_k)^T \boldsymbol{p}_\star < 0$, es decir, $\boldsymbol{p}_\star$ es una dirección de descenso para f en $\boldsymbol{x}_k$.

2. Demuestre que si el punto de Cauchy satisface $\|\boldsymbol{p}_C\|_2 < \Delta$, entonces

$$\boldsymbol{p}_C = -\left(\frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{g}_k^T B_k \boldsymbol{g}_k}\right)\boldsymbol{g}_k\,.$$

3. Sean $f(x, y) \overset{\text{def}}{=\!=} 10(x-1)^2 + y^2$, $\boldsymbol{x}_k = (2,\, 1)^T$ y $\boldsymbol{p}(s)$ definido por (1). Use Matlab u Octave para dibujar el camino de las $\boldsymbol{p}(s)$ para 100 valores equi-espaciados de $s \in [0, 50]$.

Moraleja: Las soluciones óptimas no están en una linea.

El resultado con los conjuntos de nivel puede verse como en la última página.

4. Suponga que B_k es simétrica positiva definida y que $\boldsymbol{p}(s)$ es definida por (1).

Ayuda general: Escribir $\boldsymbol{p}(s)$ *como combinación lineal de eigenvectores.*

a) Demuestre que $s_2 > s_1 \geq 0 \implies \|\boldsymbol{p}(s_2)\|_2 < \|\boldsymbol{p}(s_1)\|_2$.

Moraleja: La función $\|\boldsymbol{p}(s)\|_2$ *es estrictamente decreciente para* $s \geq 0$.

b) Suponga que $s > 0$. Demuestre que $\boldsymbol{p}(s) = \frac{\Delta}{\|\boldsymbol{p}_N\|_2}\boldsymbol{p}_N$ para $\Delta \in \big(0, \|\boldsymbol{p}_N\|_2\big)$ si y solo si $B\boldsymbol{g} = \lambda\boldsymbol{g}$ donde λ es un eigenvalor de B.

c) Sea $\boldsymbol{p}(s) \overset{\text{def}}{=\!=} -(B + sI)^{-1}\boldsymbol{g}$. Demuestre que

$$\frac{\partial}{\partial s}\boldsymbol{p}(s) = (B + sI)^{-2}\boldsymbol{g}$$

d) Sea $\boldsymbol{p}(s) \overset{\text{def}}{=\!=} -(B + sI)^{-1}\boldsymbol{g}$. Demuestre que

$$\frac{\partial}{\partial s}\|\boldsymbol{p}(s)\|_2 = -\frac{\boldsymbol{p}(s)^T(B + sI)^{-1}\boldsymbol{p}(s)}{\|\boldsymbol{p}(s)\|_2}$$

Moraleja: La función $\|\boldsymbol{p}(s)\|_2$ *es estrictamente decreciente para* $s \geq 0$.

5. Sea $\boldsymbol{p}(s)$ la curva definida por (1) con

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad , \quad \boldsymbol{g} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} .$$

Muestre que $\boldsymbol{p}(0), \boldsymbol{p}(1), \boldsymbol{p}(2)$ pertenecen al plano

$$\frac{1}{10}x - \frac{4}{5}y + z = \frac{1}{30}$$

y que $\boldsymbol{p}(3)$ no pertenece a este plano.

Moraleja: La curva de $\boldsymbol{p}(s)$ *no es plana, en general.*

6. Sea $f(x, y) \overset{\text{def}}{=\!=} \frac{1}{2}x^2 + y^2$ y $\boldsymbol{x}_0 = (1, 1)^T$. Encuentre

a) El punto de Cauchy

b) La dirección de Newton

c) La solución del (SPRC) con radio de confianza $\Delta_0 = 2$.

7. Suponga que $B \in \mathbb{R}^{n\times n}$ es simétrica positiva definida y que $\boldsymbol{g} \in \mathbb{R}^n$ es tal que $\boldsymbol{g} \neq \boldsymbol{0}$.
Usando la Factorización de Cholesky y la desigualdad de Cauchy demuestre que

$$\frac{(\boldsymbol{g}^T\boldsymbol{g})^2}{(\boldsymbol{g}^T B\boldsymbol{g})(\boldsymbol{g}^T B^{-1}\boldsymbol{g})} \leq 1$$

con igualdad si y solo si $\boldsymbol{g}$ y $B\boldsymbol{g}$ son paralelos.

8. Sea $\boldsymbol{g}_k \neq \mathbf{0}$. Resuelve el problema

$$\left[\begin{array}{ll} \text{Minimizar} & f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T \boldsymbol{p} \\ \text{Sujeto a} & \|\boldsymbol{p}\|_2 = \Delta \end{array}\right.$$

usando únicamente geometría.

9. Sea B simétrica positiva definida y $\boldsymbol{g} \neq \mathbf{0}$. Demuestre que el problema

$$\left[\begin{array}{ll} \text{Minimizar} & -\boldsymbol{g}^T \boldsymbol{p} \\ \text{Sujeto a} & \boldsymbol{p}^T B \boldsymbol{p} = 1 \end{array}\right.$$

tiene como única solución al vector

$$\boldsymbol{p}_\star = \frac{B^{-1}\boldsymbol{g}}{\sqrt{\boldsymbol{g}^T B^{-1} \boldsymbol{g}}}\,.$$

Ayuda 1: De la tarea anterior: La solución óptima con la restricción $\|\boldsymbol{p}\|_2 = 1$ *es* $\boldsymbol{g}/\|\boldsymbol{g}\|_2$.
Ayuda 2: Factorización en eigenvectores y -valores.

Soluciones

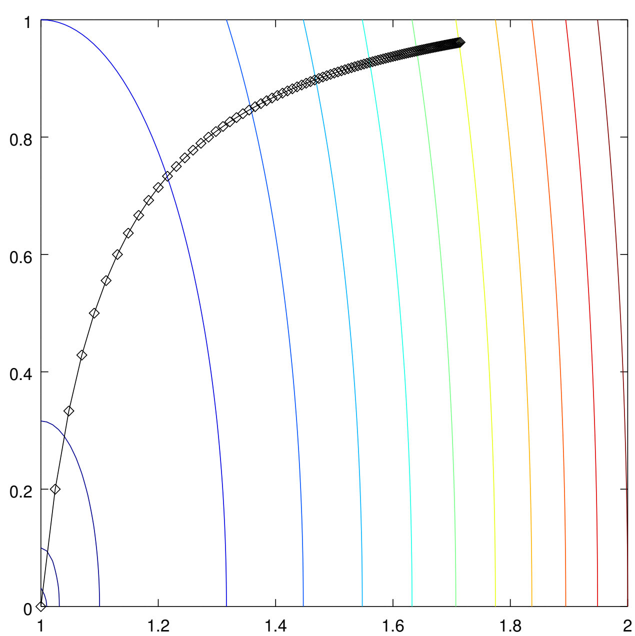

FIGURA 1. La curva $\boldsymbol{p}(s)$ para $s \in [0, 50]$ y los conjuntos de nivel de f de tarea **3.** Cada diamante indica una solución óptima a distancia $\Delta(s)$ del punto $\boldsymbol{x}_k = (2, 1)^T$.