Optimización

Primer Cuatrimestre 2022

Práctica N° 3: Direcciones conjugadas y Cuasi-Newton.

Direcciones conjugadas.

Ejercicio 1 Sea $\mathbf{Q} \in \mathbb{R}^{n\times n}$ definida positiva y sean $\mathbf{v}_1, ...\mathbf{v_n}$ vectores l.i. Mostrar que el método de Gram-Schmidt puede ser usado para generar una secuencia de direcciónes $\mathbf{Q}$-ortogonales desde los $\mathbf{v}_i$. Específicamente, muestre que

$$\mathbf{d}_1 = \mathbf{v}_1; \qquad \mathbf{d}_{k+1} = \mathbf{v}_{k+1} - \sum_{i=1}^{k} \frac{\mathbf{v}_{k+1}^t \mathbf{Q} \mathbf{d}_i}{\mathbf{d}_i^t \mathbf{Q} \mathbf{d}_i} \mathbf{d}_i$$

forma un conjunto $\mathbf{Q}-$ortogonal.

Ejercicio 2 Sea $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{Q}\mathbf{x} - \mathbf{b}^t\mathbf{x}$ con $\mathbf{Q}$ simétrica definida positiva. Sea $\mathbf{x}_1$ un minimizante de f en un subespacio S_1 que contiene al vector $\mathbf{d}$ y sea $\mathbf{x}_2$ un minimizante de f en un subespacio S_2 que contiene a $\mathbf{d}$. Mostrar que si $f(\mathbf{x}_1) < f(\mathbf{x}_2)$ entonces $\overline{\mathbf{x}} = \mathbf{x}_1 - \mathbf{x}_2$ es $\mathbf{Q}-$ortogonal a $\mathbf{d}$.

Ejercicio 3 Si definimos $\mathcal{W}_k =< \mathbf{d_0}, \cdots, \mathbf{d_{k-1}} >$ el subespacio generado por las primeras k direcciones conjugadas, mostrar que el método de las direcciones conjugadas, en cada $\mathbf{x}_k$ minimiza la función objetivo tanto en la recta $L : \mathbf{x}_{k-1} + \alpha\mathbf{d}_{k-1} : \alpha \in \mathbb{R}$, como en la variedad lineal $\mathbf{x}_0 + \mathcal{W}_k$.

Ejercicio 4 Implementar el método del Gradiente Conjugado para minimizar una función cuadrática $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{A}\mathbf{x} - \mathbf{b}^t\mathbf{x}$:

(1) A partir de un $\mathbf{x}_0$ tomar $\mathbf{d}_0 = -\mathbf{g}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

(2) Para $k = 0, 1, ...n - 1$ hacer:

  (a) $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$ con

$$\alpha_k = \frac{-\mathbf{g}_k^t\mathbf{d}_k}{\mathbf{d}_k^t\mathbf{A}\mathbf{d}_k}, \qquad \mathbf{g}_k = \mathbf{A}\mathbf{x}_k - \mathbf{b}.$$

  (b) $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k\mathbf{d}_k$ con

$$\beta_k = \frac{\mathbf{g}_{k+1}^t\mathbf{g}_{k+1}}{\mathbf{g}_k^t\mathbf{g}_k}.$$

Testear el programa con alguna función cuadrática. Considerar algún ejemplo en $\mathbb{R}^2$ y graficar los sucesivos pasos del método. Comparar con el método de máximo descenso.

Ejercicio 5 Dada una matriz simétrica definida positiva A, definimos:

$$\|\mathbf{x}\|_A^2 = \mathbf{x}^t\mathbf{A}\mathbf{x}.$$

El método de gradiente conjugado para la función cuadrática $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t\mathbf{A}\mathbf{x} - \mathbf{b}^t\mathbf{x} + c$ (o, equivalentemente, para resolver el sistema $A\mathbf{x} = b$), admite la siguiente cota de convergencia:

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_A \leq 2\left(\frac{\sqrt{\kappa(\mathbf{A})} - 1}{\sqrt{\kappa(\mathbf{A})} + 1}\right)^{k+1} \|\mathbf{x}_0 - \mathbf{x}^*\|_{\mathbf{A}},$$

donde $\kappa(\mathbf{A}) = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}$ es el número de condición de $\mathbf{A}$. Para evaluar cuán fina es esta cota, generar matrices simétricas definidas positivas con distinta distribución de autovalores, aplicar el método y graficar la velocidad de convergencia y la cota. ¿Qué se observa? ¿Depende la convergencia sólo de $\kappa(\mathbf{A})$ o incide también el modo en que están distribuidos los autovalores?

Ejercicio 6 Implementar la generalización del método de gradiente conjugado a funciones no cuadráticas, con búsqueda lineal con el criterio de Goldstein. Testearlo con las funciones de Rosenbrock, de Beale y de Ackley.

Ejercicio 7 **Tangentes paralelas.** El método de tangentes paralelas fue desarrollado originalmente para mejorar la convergencia del método del gradiente. El método consiste esencialmente en:

- Realizar una iteración del método del gradiente, obteniendo los puntos: $\mathbf{x}_0, \mathbf{x}_1$.
- Dado $\mathbf{x}_k$ calcular y_k mediante un paso del método del gradiente. Luego, tomar $\mathbf{d}_k = \mathbf{y}_k - \mathbf{x}_{k-1}$ y determinar $\mathbf{x}_{k+1}$ mediante una búsqueda lineal con dirección $\mathbf{d}_k$. La sucesión que genera el método es $\{\mathbf{x}_k\}_k$. Los puntos $\mathbf{y}_k$ se consideran valores au**x**iliares.

Esbozar un esquema gráfico del comportamiento que tendría este algoritmo. Probar que para una función cuadrática las direcciones $\{\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k\}$ son conjugadas respecto de A. Concluir que el método de tangentes paralelas es una reescritura del gradiente conjugado.

Cuasi-Newton.

Ejercicio 8 Dada una función cuadrática $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t A\mathbf{x} - \mathbf{b}^t\mathbf{x} + c$ se considera el método que toma como dirección de descenso $\mathbf{d}_k = -\mathbf{S}\mathbf{g}_k$, donde $\mathbf{g}_k = -\mathbf{r}_k = \mathbf{A}\mathbf{x}_k - \mathbf{b}$ y $\mathbf{S}$ es simétrica definida positiva. Probar que:

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_{\mathbf{A}} \leq \frac{\lambda_{\text{máx}} - \lambda_{\text{mín}}}{\lambda_{\text{máx}} + \lambda_{\text{mín}}} \|\mathbf{x}_k - \mathbf{x}^*\|_{\mathbf{A}},$$

donde $\lambda_{\text{máx}}$ y $\lambda_{\text{mín}}$ son los autovalores máximo y mínimo de $\mathbf{SA}$. **Sug.:** realizar un un cambio de variables y una redefinición de matrices y recordar que dos matrices semejantes comparten autovalores.

Ejercicio 9 Implementar el método DFP (Davidon, Fletcher y Powell).

Ejercicio 10 Mostrar que si se toma $\mathbf{H}_0 = \mathbf{I}$ el método DFP es el método de gradiente conjugado.

Ejercicio 11 Implementar el método BFGS (Broyden, Fletcher, Goldfarb, Shanno)

Ejercicio 12 Resolver el problema de la braquistocrona con los métodos de Newton y BFSG. Para Newton, utilizar dos versiones: una que utilice el gradiente y el hessiano analíticos, y una que los estime mediante diferencias finitas. Utilizar como dato inicial los valores de $\mathbf{y}$ en la recta $y = 1 - \frac{2x}{\pi}$, que corresponde a la caída de la partícula por un plano inclinado. La solución óptima puede parametrizarse por:

$$\begin{aligned} x(t) &= \frac{1}{2}(t - \sin(t)), \\ y(t) &= 1 - \frac{1}{2}(1 - \cos(t)), \end{aligned}$$

para $t \in [0, \pi]$. Siendo t es una variable puramente matemática, sin significado físico. Estimar la velocidad de convergencia de cada método comparando con esta solución.