

Motivaciones. Aproximamos al gradiente y a la matriz Hessiana de una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ utilizando diferencias finitas. Los pasos que propongo son empíricamente cerca del paso óptimo, ver Matemática Computacional o el documento `addInfo_FD.pdf` en comunidad.itam.

1. TEORÍA

¿Cómo aproximar la primera derivada en una dimensión?

Para $f: \mathbb{R} \rightarrow \mathbb{R}$ (diferenciable) usamos la aproximación de matemática computacional:

$$(1) \quad f'(a) \approx \frac{f(a+h) - f(a-h)}{2h} \quad \text{con} \quad h \stackrel{\text{def}}{=} (\varepsilon_M)^{1/3} (|a| + 1),$$

donde $\varepsilon_M \approx 2 \cdot 10^{-16}$ es el épsilon de la máquina. De la teoría de Matemática Computacional se sabe que este h es (empíricamente) cerca del paso óptimo.

¿Cómo aproximar el gradiente de $f: \mathbb{R}^n \rightarrow \mathbb{R}$?

Tenemos que formar la diferencia (1) para cada entrada x_i del vector $\mathbf{x} \in \mathbb{R}^n$, es decir, si \mathbf{e}_i es el vector canónico (con 1 en la entrada i), entonces

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \approx \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x} - h\mathbf{e}_i)}{2h}.$$

En la definición de h debemos usar $|a| \stackrel{\text{def}}{=} |x_i|$.

¿Cómo aproximar la matriz Hessiana de $f: \mathbb{R}^n \rightarrow \mathbb{R}$?

Consideramos $n = 2$ y $f(x, y)$ primero. Sea $g(y) = \frac{\partial f}{\partial x}(x, y)$. Usando la aproximación de arriba tenemos

$$\begin{aligned} \frac{\partial^2}{\partial y \partial x} f(x, y) &\stackrel{\text{def}}{=} \frac{\partial}{\partial y} g(y) \approx \frac{1}{2h} [g(y+h) - g(y-h)] \\ &\stackrel{\text{def}}{=} \frac{1}{2h} \left[\frac{\partial f}{\partial x}(x, y+h) - \frac{\partial f}{\partial x}(x, y-h) \right] \\ &\approx \frac{1}{4h^2} \left[(f(x+h, y+h) - f(x-h, y+h) - f(x+h, y-h) + f(x-h, y-h)) \right]. \end{aligned}$$

El caso general consideramos $x = x_i, y = x_j$ (en esa formula) y usamos vectores canónicos para obtener:

$$\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{x}) \approx \frac{1}{4h^2} \left[(f(\mathbf{x} + h\mathbf{e}_i + h\mathbf{e}_j) - f(\mathbf{x} - h\mathbf{e}_i + h\mathbf{e}_j) - f(\mathbf{x} + h\mathbf{e}_i - h\mathbf{e}_j) + f(\mathbf{x} - h\mathbf{e}_i - h\mathbf{e}_j)) \right].$$

De la diferencia centrada en una dimensión se sabe que $h \stackrel{\text{def}}{=} (\varepsilon_M)^{1/4} (|a| + 1)$ es cerca del paso óptimo. Para usarlo con la diferencia mixta, escogemos $h_i \mathbf{e}_i \stackrel{\text{def}}{=} (\varepsilon_M)^{1/4} (|x_i| + 1) \mathbf{e}_i$.

2. PRÁCTICA

1. Utilizando la teoría de la página anterior escriba una función en MatLab (Octave) llamada `apGrad.m` que tiene dos argumentos (una función anónima f y un vector \mathbf{x}). El resultado de `apGrad(f, x)` debe ser una aproximación del gradiente $\nabla f(\mathbf{x})$.
2. Utilizando la teoría de la página anterior escriba una función en MatLab (Octave) llamada `apHess.m` que tiene dos argumentos (una función anónima f y un vector \mathbf{x}). El resultado de `apHess(f, x)` debe ser una aproximación de la Hessiana $\nabla^2 f(\mathbf{x})$.
3. Sea f la función cuadrática definida por $f(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + 1$, con

$$A \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}, \quad \mathbf{b} \stackrel{\text{def}}{=} \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}.$$

La función f tiene el único mínimo en $\mathbf{x}^* = (1, 0, 0, 0)^T$.

Escriba una función, llamada `fPascal.m` sin argumentos que devuelve como respuesta las funciones anónimas f , el gradiente exacto ∇f y la Hessiana exacta $\nabla^2 f$.

Ayuda 1: La instrucción de MatLab llamada `pascal` te da la matriz.

Ayuda 2: Demuestre $\nabla f(\mathbf{x}) = \frac{1}{2}(A + A^T)\mathbf{x} + \mathbf{b}$ y $\nabla^2 f(\mathbf{x}) = \frac{1}{2}(A + A^T)$.

4. Sea $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ la función de Rosenbrock:

$$f(\mathbf{x}) \stackrel{\text{def}}{=} 100(x_1^2 - x_2)^2 + (x_1 - 1)^2.$$

Escriba una función, llamada `fRosenbrock2d.m` sin argumentos que devuelve como respuesta las funciones anónimas f , el gradiente exacto ∇f y la Hessiana exacta $\nabla^2 f$.

5. (*Calidad de aproximaciones*) Para la función cuadrática arriba se pide

- aproximar el gradiente en el punto $(4, 4, 4, 4)^T$.
- evaluar el gradiente exacto en el punto $(4, 4, 4, 4)^T$.
- medir el error absoluto en la norma $\|\cdot\|_\infty$.
- medir el máximo de los errores relativos en cada entrada del gradiente.

Repita los primeros 2 incisos para el punto óptimo $(1, 0, 0, 0)^T$.

6. Repita el ejercicio anterior con la función de Rosenbrock en el punto $(2, 3)^T$ y en el óptimo $(1, 1)^T$.
7. Repita los ejercicios 5. y 6. para las Hessianas en los mismos puntos.
8. *Eficiencia.* En Matlab (Octave) se puede medir el tiempo con las instrucciones `tic` y `toc`. Comparé el tiempo de evaluación. Por ejemplo, mide el tiempo de aproximar 1000 veces el gradiente, y evaluar 1000 veces el gradiente exacto. Después, comparé los tiempos.