Motivaciones. Aproximamos al gradiente y a la matriz Hessiana de una función $f\colon \mathbb{R}^n \to \mathbb{R}$ 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 $f\colon \mathbb{R} \to \mathbb{R}$ (diferenciable) usamos la aproximación de matemática computacional:

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

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 $f\colon \mathbb{R}^n \to \mathbb{R}$?

Tenemos que formar la diferencia (1) para cada entrada x_i del vector $\boldsymbol{x} \in \mathbb{R}^n$, es decir, si $\boldsymbol{e}_i$ es el vector canónico (con 1 en la entrada i) y $\boldsymbol{h}_i \stackrel{\text{def}}{=\!=} \big(\varepsilon_M\big)^{1/3}(|x_i|+1)\boldsymbol{e}_i,$ entonces la definición en una dimensión tiene la forma general:

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) \approx \frac{f(\boldsymbol{x}+\boldsymbol{h}_i) - f(\boldsymbol{x}-\boldsymbol{h}_i)}{2\,\|\boldsymbol{h}_i\|}\,.$$

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

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

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

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

$$\frac{\partial^2}{\partial x_j \partial x_i} f(\boldsymbol{x}) \approx \frac{1}{4\,\|\boldsymbol{h}_i\|\,\|\boldsymbol{h}_j\|}\Big[(f(\boldsymbol{x}+\boldsymbol{h}_i+\boldsymbol{h}_j) - f(\boldsymbol{x}-\boldsymbol{h}_i+\boldsymbol{h}_j) - f(\boldsymbol{x}+\boldsymbol{h}_i-\boldsymbol{h}_j) + f(\boldsymbol{x}-\boldsymbol{h}_i-\boldsymbol{h}_j)\Big]\,.$$

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 $\boldsymbol{h}_i \stackrel{\text{def}}{=\!=} (\varepsilon_M)^{1/4}(|x_i|+1)\boldsymbol{e}_i \in \mathbb{R}^n,$ similar al paso usado para aproximar el gradiente.

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 $\boldsymbol{x}$). El resultado de `apGrad(f, x)` debe ser una aproximación del vector $\nabla f(\boldsymbol{x})$.

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 $\boldsymbol{x}$). El resultado de `apHess(f, x)` debe ser una aproximación de la Hessiana $\nabla^2 f(\boldsymbol{x})$.

3. Sea f la función cuadrática definida por $f(\boldsymbol{x}) \stackrel{\text{def}}{=\!=} \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + 1$, con

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

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

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

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

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

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

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

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.