Optimización Tarea 5

Francisco Javier Peralta Ramírez

27 de marzo de 2018

1. Implemente el método del gradiente conjugado lineal
2. Con la implementación anterior, resuelva el siguiente problema (Smoothing model).

$$f(x) = \sum_{i=1}^{n}(x_i - y_i)^2 + \lambda \sum_{i=1}^{n-1}(x_{i+1} - x_i)^2$$

Para $\lambda \in 1, 100, 1000$

Calculamos el gradiente y el Hessiano:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2(x_1 - y_1) - 2\lambda(x_2 - x_1) \\ 2(x_2 - y_2) + 2\lambda(2x_2 - x_1 - x_3) \\ \vdots \\ 2(x_n - y_n) + 2\lambda(x_n - x_{n-1}) \end{pmatrix}$$

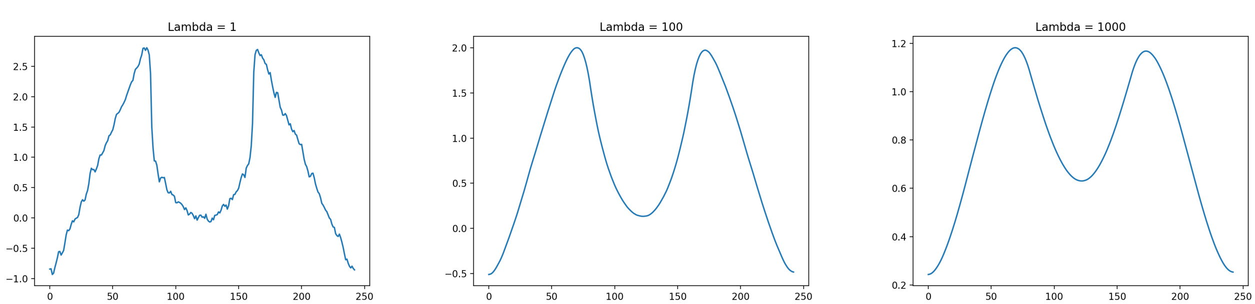
$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 2+2\lambda & -2\lambda & 0 & \cdots & 0 \\ -2\lambda & 2+4\lambda & \ddots & \cdots & \vdots \\ 0 & \cdots & \ddots & \ddots & -2\lambda \\ 0 & \cdots & 0 & -2\lambda & 2+2\lambda \end{pmatrix}$$

Esta función es de suavizado, donde tenemos un vector de $\boldsymbol{x}$ que queremos hacer similar a $\boldsymbol{y}$, no queremos interpolar ya que la variable $\boldsymbol{y}$ tiene ruido, por lo que penalizamos los cambios en $\boldsymbol{x}$ con un cierto peso λ, entre más grande el peso más suave la curva.

Para poder usar el gradiente conjugado, tenemos que ver el problema como uno de la forma $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$. Donde Q es una matriz simétrica y definida positiva. Si tomamos la expansión de Taylor:

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) = (\boldsymbol{x} - \boldsymbol{a})\nabla f(\boldsymbol{a}) + (\boldsymbol{x} - \boldsymbol{a})H(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{a})$$

Podemos tomar $\boldsymbol{a} = 0$ y $\nabla f(\boldsymbol{a}) = -b$ obtendríamos la función $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T H \boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + f(\boldsymbol{0})$ el término extra $f(\boldsymbol{0}$ se elimina al derivar, así que nuestro problema a resolver sigue siendo $Q\boldsymbol{x} = \boldsymbol{b}$.



(a) $\lambda = 1$ (b) $\lambda = 100$ (c) $\lambda = 1000$

3. Implemente los métodos de gradiente conjugado no lineal: Fletcher-Reeves, Polak-Ribiere y Hestenes-Stiefel.

   Notamos que el algoritmo es el mismo para todos los métodos anteriores, con la diferencia en el cálculo de la β, por lo que podemos crear un algoritmo principal que y sólo pasar la función para calcular la β como parámetro del algoritmo. Para calcular el α podemos usar *backtracking*, así eliminando la necesidad de calcular el *Hessiano*.

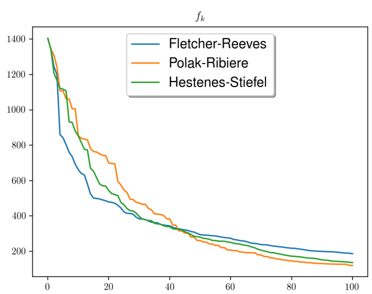
   Aplicar dichos métodos para resolver los siguientes problemas:

   a) Rosembrock, n = 100

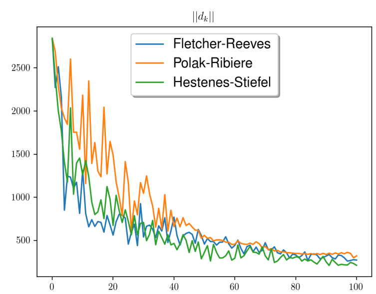
$$
\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=0}^{n-1}[100(x_{i+1}-x_i^2)^2+(1-x_i)^2] \\
\boldsymbol{x^0} &= [-1.2, 1, 1, \cdots, 1, -1.2, 1]^T \\
\boldsymbol{x^*} &= [1, 1, \cdots, 1, 1]^T \\
f(\boldsymbol{x^*}) &= 0
\end{aligned}
$$

   Calculamos el gradiente:

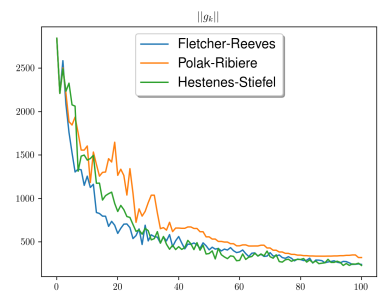
$$
\nabla f(\boldsymbol{x}) = \begin{pmatrix} -400(x_2-x_1^2)x_1 - 2(1-x_1) \\ 200(x_2-x_1^2) - 400(x_3-x_2^2)x_2 - 2(1-x_2) \\ \vdots \\ 200(x_i-x_{i-1}^2) - 400(x_{i+1}-x_i^2)x_i - 2(1-x_i) \\ \vdots \\ 200(x_n - x_{n-1}^2) \end{pmatrix}
$$



(a) $f(\boldsymbol{x}_k)$



(b) $||\boldsymbol{d}_k||$



(c) $||\boldsymbol{g}_k||$

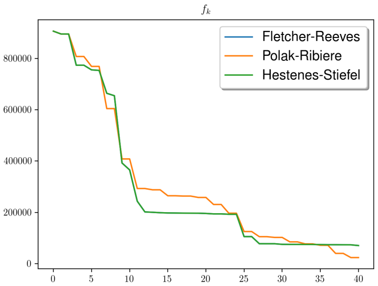
   b) Wood

$$
\begin{aligned}
f(\boldsymbol{x}) &= 100(x_1^2-x_2)^2+(x_1-1)^2+(x_3-1)^2+90(x_3^2-x_4)^2 \\
&\quad + 10.1[(x_2-1)^2+(x_4-1)^2] + 19.8(x_2-1)(x_4-1) \\
\boldsymbol{x^0} &= [-3, -1, -3, -1]^T \\
\boldsymbol{x^*} &= [1, 1, 1, 1]^T \\
f(\boldsymbol{x^*}) &= 0
\end{aligned}
$$

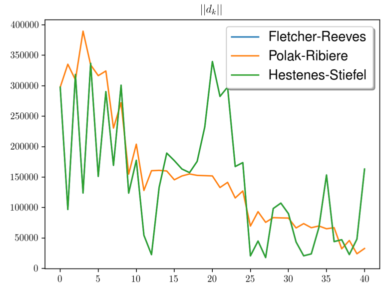
   Calculamos el gradiente

$$
\nabla f(\boldsymbol{x}) = \begin{pmatrix} 400(x_1^2-x_2)x_1 + 2(x_1-1) \\ -200(x_1^2-x_2) + 20.2(x_2-1) + 19.8(x_4-1) \\ 2(x_3-1) + 360(x_3^2-x_4)x_3 \\ -180(x_3^2-x_4) + 20.2(x_4-1) + 19.8(x_2-1) \end{pmatrix}
$$

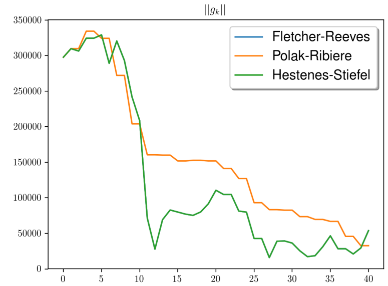
   En este caso las gráficas para Flether-Reeves y Hestenes se enciman.



(a) $f(\boldsymbol{x}_k)$



(b) $||\boldsymbol{d}_k||$



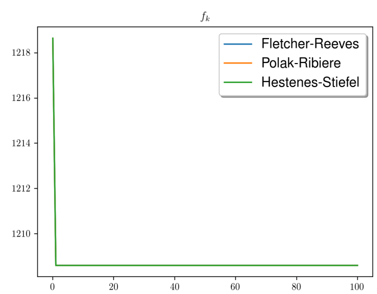
(c) $||\boldsymbol{g}_k||$

c) Convex 1

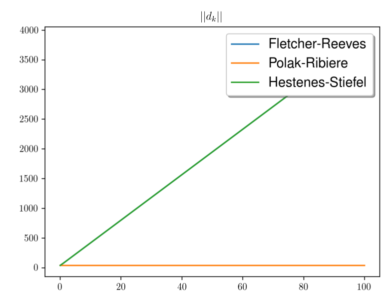
$$
\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{n} e^{x_i} - x_i \\
\boldsymbol{x^0} &= [\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-11}{n}, 1]^T \\
\boldsymbol{x^*} &= [0, 0, \cdots, 0, 0]^T \\
f(\boldsymbol{x^*}) &= n
\end{aligned}
$$

Calculamos el gradiente

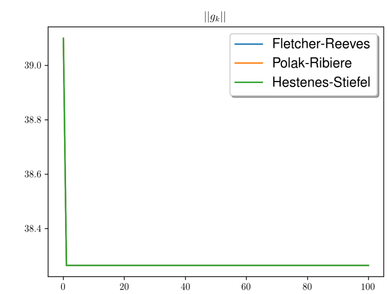
$$
\nabla f(\boldsymbol{x}) = \begin{pmatrix} e^{x_1} - 1 \\ e^{x_2} - 1 \\ \vdots \\ e^{x_{n-1}} - 1 \\ e^{x_n} - 1 \end{pmatrix}
$$



(a) $f(\boldsymbol{x}_k)$



(b) $||\boldsymbol{d}_k||$



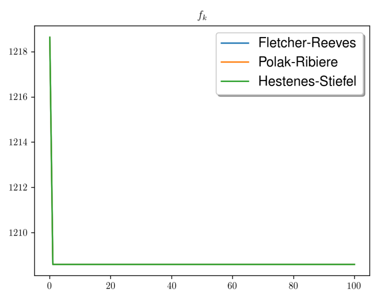
(c) $||\boldsymbol{g}_k||$

d) Convex 2

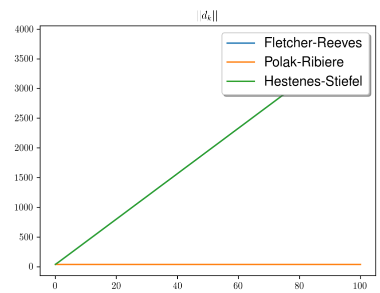
$$
\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{n} \frac{i}{n} e^{x_i} - x_i \\
\boldsymbol{x^0} &= [1, 1, \cdots, 1, 1]^T \\
\boldsymbol{x^*} &= [0, 0, \cdots, 0, 0]^T \\
f(\boldsymbol{x^*}) &= \frac{n+1}{2}
\end{aligned}
$$

Calculamos el gradiente

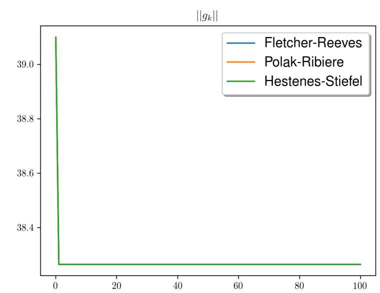
$$
\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{1}{n}(e^{x_1} - 1) \\ \frac{2}{n}(e^{x_2} - 1) \\ \vdots \\ \frac{n-1}{n}(e^{x_{n-1}} - 1) \\ \frac{n}{n}(e^{x_n} - 1) \end{pmatrix}
$$



(a) $f(\boldsymbol{x}_k)$



(b) $||\boldsymbol{d}_k||$



(c) $||\boldsymbol{g}_k||$