MATH520 Homework 3

Fernando

February 13, 2024

Exercise 8.18

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T b$, where $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{Q}$ is a real symmetric positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$. Show that the method converges in one step, that is, $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b}$, if and only if $\boldsymbol{x}^{(0)}$ is chosen such that $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ is an eigenvector of $\boldsymbol{Q}$.

Solution

$(\implies)$

For this part we don't need $\boldsymbol{Q}$ to be positive definite, invertible is enough.

First we note that $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}$. And thus

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha_0 \nabla f(\boldsymbol{x}^{(0)}) = \boldsymbol{x}^{(0)} - \alpha_0(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}),$$

Where $\alpha_0 \neq 0$ because $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$. Then

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{Q}^{-1}\boldsymbol{b} \\
\boldsymbol{x}^{(0)} - \alpha_0(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}) &= \boldsymbol{Q}^{-1}\boldsymbol{b} \\
\boldsymbol{Q}(\boldsymbol{x}^{(0)} - \alpha_0(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b})) &= \boldsymbol{b} \\
\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b})) &= \frac{1}{\alpha_0}(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}).
\end{aligned}$$

$(\impliedby)$

For this part we will use the positive definiteness of $\boldsymbol{Q}$.

$$\begin{aligned}
\boldsymbol{Q}(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b})) &= \lambda(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}) \\
\boldsymbol{Q}(\boldsymbol{x}^{(0)} - \frac{1}{\lambda}(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b})) &= \boldsymbol{b} \\
\boldsymbol{x}^{(0)} - \frac{1}{\lambda}(\boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}) &= \boldsymbol{Q}^{-1}\boldsymbol{b}
\end{aligned}$$

If we can proof that $\frac{1}{\lambda}$ minimizes $\varphi_0(\alpha) = f(\boldsymbol{x}^{(0)} - \alpha\nabla f(\boldsymbol{x}^{(0)}))$ then the left hand side of the last equation is exactly $\boldsymbol{x}^{(1)}$ and we are done.

Luckily in this case we know the minimizer, which according to the book is

$$\frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{(\boldsymbol{g}^{(0)})^T\boldsymbol{Q}\boldsymbol{g}^{(0)}}$$

Notice that in our case $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ and because by hypothesis it is an eigenvector we have:

$$\begin{aligned}
\boldsymbol{Q}\boldsymbol{g}^{(0)} &= \lambda\boldsymbol{g}^{(0)} \\
(\boldsymbol{g}^{(0)})^T\boldsymbol{Q}\boldsymbol{g} &= \lambda(\boldsymbol{g}^{(0)})^T\boldsymbol{g} \\
\frac{1}{\lambda} &= \frac{(\boldsymbol{g}^{(0)})^T\boldsymbol{g}^{(0)}}{(\boldsymbol{g}^{(0)})^T\boldsymbol{Q}\boldsymbol{g}^{(0)}}
\end{aligned}$$

so $\frac{1}{\lambda}$ in fact minimizes $\varphi_0(\alpha)$ and we are done.

Exercise 8.24

Given $f : \mathbb{R}^n \to \mathbb{R}$, consider the general iterative algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)},$$

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots$ are given vectors in $\mathbb{R}^n$ and α_k is chosen to minimize $f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}).$$

Show that for each k, the vector $\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$ is orthogonal to $\nabla f(\boldsymbol{x}^{(k+1)})$ (assuming that the gradient exists).

Solution

Let us define the function

$$h(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$$

then

$$h'(\alpha) = \nabla f(x^{(k)} + \alpha\boldsymbol{d}^{(k)}) \cdot \boldsymbol{d}^{(k)}.$$

By definition α_k satisfies $h'(\alpha_k) = 0$ (FONC) so

$$0 = h'(\alpha_k) = \nabla f(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) \cdot \boldsymbol{d}^{(k)} = \nabla f(\boldsymbol{x}^{(k+1)}) \cdot (\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$$

Which is what we wanted to prove.

Exercise 9.1

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = (x - x_0)^4$, where $x_0 \in \mathbb{R}$ is constant. Suppose that we apply Newton's method to the problem of minimizing f, with iterates $x^{(0)}, x^{(1)}, x^{(2)}, \ldots$

a. Write down the update equation for Newton's method applied to the problem.

b. Let $y^{(k)} = |x^{(k)} - x_0|$, where $x^{(k)}$ is the kth iterate in Newton's method. Show that the sequence $\{y^{(k)}\}$ satisfies $y^{(k+1)} = \frac{2}{3} y^{(k)}$.

c. Show that $x^{(k)} \to x_0$ for any initial guess $x^{(0)}$.

d. Show that the order of convergence of the sequence $\{x^{(k)}\}$ in part b is 1.

e. Theorem 9.1 states that under certain conditions, the order of convergence of Newton's method is at least 2. Why does that theorem not hold in this particular problem?

Solution a

in this case

$$\boldsymbol{F}(x) = 12(x - x_0)^2$$

so

$$\boldsymbol{F}(x^{(k)})^{-1} = \frac{1}{12(x^{(k)} - x_0)^2}$$

and

$$\boldsymbol{g}^{(k)} = 4(x^{(k)} - x_0)^3$$

so

$$x^{(k+1)} = x^{(k)} - \frac{4(x^{(k)} - x_0)^3}{12(x^{(k)} - x_0)^2} = x^{(k)} - \frac{x^{(k)} - x_0}{3} = \frac{2x^{(k)} + x_0}{3}.$$

Solution b

$$y^{(k+1)} = |x^{(k+1)} - x_0| = \left| \frac{2x^{(k)} + x_0}{3} - x_0 \right| = \left| \frac{2x^{(k)} - 2x_0}{3} \right| = \frac{2}{3} y^{(k)}.$$

Solution c

This is equivalent to proving that $y^{(k)} \to 0$. From b we see that an explicit formula for $y^{(k)}$ is $y^{(k)} = \left(\frac{2}{3}\right)^k y^{(0)}$. Taking the limit as $k \to 0$ we get the result. Notice that the value of $y^{(0)}$ is irrelevant.

Solution d

We notice that

$$\lim_{k\to\infty} \frac{|x^{(k+1)} - x_0|}{|x^{(k)} - x_0|^p} = \lim_{k\to\infty} \frac{y^{(k+1)}}{\left(y^{(k)}\right)^p} = \frac{2}{3} \lim_{k\to\infty} \frac{y^{(k)}}{(y^{(k)})^p}$$

$$= \frac{2}{3} \lim_{k\to\infty} (y^{(k)})^{1-p} = \frac{2}{3} (y^{(0)})^{1-p} \lim_{k\to\infty} \left(\left(\frac{2}{3}\right)^{(1-p)} \right)^k .$$

Then we see that the biggest value p can take is 1 because any p bigger than that would cause the limit to become infinite. With $p = 1$ the limit is 1.

Solution e

In this case $F(x^*) = 0$ so $F(x^*)$ is not invertible and the theorem doesn't hold.

Exercise 9.4

Consider Rosenbrock's function: $f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, where $\boldsymbol{x} = [x_1, x_2]^T$. This function is also known as the banana function because of the shape of its level sets.

a. Prove that $[1, 1]^T$ is the unique global minimizer of f over $\mathbb{R}^2$.

b. With a starting point of $[0, 0]^T$, apply two iterations of Newton's method (with unit step size).

c. Repeat part b using a gradient algorithm with a fixed step size of $\alpha_k = 0.05$ at each iteration.

Solution a

In this case

$$\nabla f(\boldsymbol{x}) = [-400(x_2 - x_1^2)x_1 + 2(x_1 - 1), 200(x_2 - x_1^2)]^T$$

and we can see that the only point that satisfies $\nabla f = 0$ is $[1, 1]^T$. Also we have:

$$\boldsymbol{F}(x) = \begin{bmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

and then the matrix $F([1, 1]^T)$ is positive definite (I used WolframAlpha to compute the eigenvalues) so $[1, 1]^T$ is a unique minimizer.

Solution b

$$\boldsymbol{x}^{(1)} = [0, 0]^T - \frac{1}{200} \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = [1, 0]^T$$

$$\boldsymbol{x}^{(2)} = [1, 0]^T - \frac{1}{40200} \begin{bmatrix} 100 & 200 \\ 200 & 601 \end{bmatrix} \begin{bmatrix} 400 \\ -200 \end{bmatrix} = [1, 1]^T.$$

Solution c

$$\boldsymbol{x}^{(1)} = [0,0]^T - 0.05 \begin{bmatrix} -2 \\ 0 \end{bmatrix} = [0.1, 0]^T$$

$$\boldsymbol{x}^{(2)} = [0.1,0]^T - 0.05 \begin{bmatrix} -1.4 \\ -2 \end{bmatrix} = [0.17, 0.1]^T.$$